UNITED STATES DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233-0001

May 19, 2022

DSSD 2020 POST-ENUMERATION SURVEY MEMORANDUM SERIES #2020-J-05

MEMORANDUM FOR	Timothy L. Kennel Assistant Division Chief, Statistical Methods Decennial Statistical Studies Division
Prepared by:	Michael Beaghen (*signed*) Richard N. Turner (*signed*) Mark Jost (*signed*) Elizabeth Marra (*signed*) Decennial Statistical Studies Division
Subject:	2020 Post-Enumeration Survey Estimation Methods: Missing Data for Person Estimates

This report is part of a series describing the methodology and results of the 2020 Post-Enumeration Survey (PES) estimation activities. Specifically, this report summarizes the 2020 PES methods for accounting for missing data and results of these methods. The missing data methods include noninterview adjustment, imputation for match status, and imputation for correct enumeration status.

For more information, contact Michael Beaghen at Michael.a.beaghen@census.gov or at 301-763-2981, or Elizabeth Marra at Elizabeth.marra@census.gov or at 301-763-7760.

cc:
DCMD CMDE List

[This page is intentionally left blank]

2020 Post-Enumeration Survey

Estimation Methods: Missing Data for Person Estimates

Coverage Measurement Design and Estimation (CMDE)
Integrated Project Team (IPT)

IPT Lead: Timothy Kennel

Report Authors: Michael Beaghen, Richard N. Turner, Mark Jost, Elizabeth Marra

Version 1.0

May 19, 2022

Table of Contents

List of Tables

Executive Summary

The Post-Enumeration Survey (PES) estimated the net coverage error and the components of coverage of the 2020 Census. Like all surveys, the PES had to deal with missing or incomplete response data. Some respondents did not answer specific questions needed to estimate the population size or components of coverage. When this happened, we imputed values to fill in the missing data. Further, some households refused to respond to the survey or did not provide enough information to support requirements for the PES, in which case we performed a noninterview adjustment. This report describes the methodology and results of the processes the PES undertook to treat missing data.

<u>Missing Data Needed to Estimate the Population Size</u>

The estimation of population size, and thus net coverage error, used both the P sample, an independent PES listing of housing units and people in those housing units, and the E sample, a sample of census housing unit and person enumerations.

- The levels of missing data in the 2020 PES were higher than those of the previous post-enumeration survey.

    - *Noninterviews in the P sample.* Among occupied housing units, the interview rate was 83.2 percent. This was lower than the interview rate of 96.6 percent in the 2010 Census Coverage Measurement (CCM) survey.
    - *Missing enumeration status in the Person E sample.* We did not have enough information to determine if 11.6 percent of the person enumerations in the E sample were correct or erroneous enumerations. The unresolved enumeration status rate for the 2010 CCM was 4.8 percent.
    - *Missing inclusion status in the Person P sample.* We did not have enough information to determine if 6.98 percent of the people listed in the PES met the definition of being in-scope for the P sample. The unresolved inclusion status rate for the 2010 CCM was 2.87 percent.
    - *Missing match status in the Person P sample.* We did not have enough information to determine if 5.02 percent of P-sample people matched to enumerations in the 2020 Census. The unresolved match status rate for the 2010 CCM was 1.90 percent.

- Weighting adjustments and statistical imputation were used to mitigate the effects of missing data in both the P and E samples. The correct enumeration and match rates were not expected to be the same before and after imputation, because cases with missing data were not necessarily like those with complete data.

- *Imputation of correct enumeration status.* The correct enumeration rate before imputation was 86.75 percent. After the imputation of unresolved E-sample enumerations, it was 87.16 percent.
- *Imputation of match rate.* The P-sample match rate before imputation was 86.77 percent. After the imputation of unresolved P-sample people, it was 84.98 percent.

Missing Data Needed to Estimate Components of Census Coverage

To support the estimates of components of census coverage, the PES also classified E-sample enumerations into one of several enumeration statuses instead of the binary outcome of correct or erroneous census enumeration.

- *Missing data for enumeration statuses.* The amount of missing data for the enumeration statuses used to estimate components of census coverage in the 2020 PES was higher than in the 2010 CCM. About 12 percent of the E-sample enumerations were unresolved for most of the enumeration statuses used to calculate components of census coverage.

- *Unresolved duplication status.* The unresolved rate for erroneous enumeration because of duplication was lower than the unresolved rate for other components of census coverage. About 0.4 percent of all E-sample enumerations were unresolved with respect to duplication.

1. Introduction

This document provides an overview of how missing data were handled in the 2020 Post-Enumeration Survey (PES)[1] for the United States person estimates. It describes the missing data procedures used to support the estimation of net coverage error and the components of census coverage for people and the results of these procedures. It focuses on the noninterview adjustment and on the imputation of match and correct enumeration statuses used in the estimation. It does not discuss characteristic imputation for PES housing units or people (refer to Phan and Lawrence, 2022).

Documentation and results for the housing unit estimation and for Puerto Rico will appear in future reports. Documentation of the overall PES design can be found in Kennel (2019).

1.1 Post-Enumeration Survey Missing Data

The PES used dual-system estimation to estimate the population size of the nation. By comparing the PES population estimate to the 2020 Census, the PES estimated the net coverage error of the 2020 Census count of people. The dual-system estimator used by the 2020 PES required both a probability of match and a probability of correct enumeration; refer to Zamora (2022) for details on how the dual-system estimates were calculated.

The PES consisted of two samples: a sample of the population or P sample, and a sample of census enumerations or E sample. The P sample of housing units and people in housing units was enumerated independently of the census. The E sample consisted of census housing units and census person enumerations in housing units in the same sample areas as the P sample.

We used the P sample to estimate the match probability and the E sample to estimate the correct enumeration probability. The enumeration status indicated whether a census person enumeration should have been counted in the sample block search area[2] on Census Day. Before calculating dual-system estimates and estimates of the components of census coverage, we had to account for missing data in the P and E samples. We encountered three types of missing data in the PES samples.

1. *Household-level noninterviews in the person P sample.* For some of these noninterviews, the household could not be contacted or the interview was refused or not completed. However, more commonly, the information provided by the respondent was not

[1] The Census Bureau's Disclosure Review Board has reviewed this product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release. DRB number CBDRB-FY22-244.

[2] To be more precise, it was not the 'block search area' but the 'basic collection unit search area.' A basic collection unit was the smallest geographic level for 2020 Census data collection and roughly corresponded to a block. Refer to Hogan (2003) for more details on the search area.

complete enough to accurately match anyone in the household to the census. The noninterview adjustment spread the weights of household noninterviews to similar households that were interviewed.

2. *Unresolved statuses in the P and E samples.* When we refer to status, we usually mean the answer to a question we needed to estimate coverage. There are four statuses discussed in this memo:
    - E-sample enumeration status: Was a person correctly enumerated or erroneously enumerated?
    - P-sample inclusion status: Did a person meet the requirements for being in-scope for the PES?
    - P-sample mover status: Did a person move between April 1, 2020 and the PES interview?
    - P-sample match status: Was a person in the PES correctly counted in the 2020 Census?

   The statuses provided the information needed to calculate dual-system estimates and the estimates of the components of census coverage. Missing statuses arose when we did not have enough information about a person to make a confident determination. When a status was missing, we imputed a probability for that status using information available about the person and about resolved cases with similar characteristics.

3. *Missing demographic characteristics in the P sample*. This situation occurred when a person was missing age, sex, relationship, tenure, race, or Hispanic origin. The characteristic imputation methods are discussed in Phan and Lawrence (2022). We do not discuss them in this report.

1.2 Preliminaries: Sufficient Information for Dual-System Estimation and Whole-Person Imputations

Several concepts are important throughout this report and are defined here. The first is sufficient information for dual-system estimation. Person records with sufficient information for dual-system estimation had adequate information to uniquely identify an individual. For example, a first and last name are needed to uniquely identify a person and certainly needed to accurately determine if someone in the PES matched to a census record. In contrast, person records with insufficient information for dual-system estimation did not meet the minimum threshold to uniquely identify a person. We could not determine with confidence the inclusion, match, or enumeration statuses of insufficient information cases using the PES matching and field operations. For many of the insufficient information cases, a full name was missing. For simplicity, we use the terms “sufficient information” and “insufficient information” throughout the remainder of this document.

Table 1 and Table 2 show for the P and E samples, respectively, the percentage of insufficient information cases in the 2020 PES and the 2010 CCM[3]. We note higher rates of insufficient information for both the P and E samples in the 2020 PES.

Table 1: P-Sample Insufficient Information Counts and Rates

	2020			2010		
	Total	Insufficient Information Count	Insufficient Information (Percent)	Total	Insufficient Information Count	Insufficient Information (Percent)
All PES Listed Cases	345,000	51,000	14.8	393,000	13,000	3.3
Post-NIA P-sample Cases	301,000	12,500	4.2	383,000	6,400	1.7

Note: NIA stands for Noninterview Adjustment.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

Through interviews that were independent of the 2020 Census, the PES listed 345,000 people. Of those people, 51,000 did not have sufficient information for dual-system estimation. If everyone in the household had insufficient information, the household was called a noninterview and its weight distributed among responding households in a process called the noninterview adjustment (NIA); refer to Section 2 for details on this process. After the NIA, there were still 12,500 people who had insufficient information, but they were in households where at least one person had sufficient information. These had their inclusion and match statuses imputed (refer to Section 5 for details). Note that there were also person listings with sufficient information whose weights were distributed in the NIA process, or were removed from PES processing because of PES data editing rules.

Table 2: E-Sample Insufficient Information Counts and Rates

2020			2010		
Total	Insufficient Information Count	Insufficient Information (Percent)	Total	Insufficient Information Count	Insufficient Information (Percent)
397,000	40,000	10.1	384,000	13,000	3.4

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

We treated insufficient information E-sample cases as erroneous for the estimation of net coverage error, but attempted to match them and determine their enumeration status for the estimation of components of coverage. If we could not match an E-sample enumeration with insufficient information, we imputed its enumeration statuses for the estimation of components of coverage since we could not send it to a follow-up interview.

[3] The post-enumeration survey in 2010 was called the Census Coverage Measurement survey.

Another important concept is whole-person imputation. A whole-person census imputation was a census record for which all the person characteristics were imputed. This occurred when very little information about the household or the people in the household was obtained. There were about 10,850,000 whole-person imputations in the 2020 Census (Khubba et al., 2022). In contrast, an enumeration with insufficient information had some reported person characteristics.

2. Noninterview Adjustment

The PES estimation included a NIA to mitigate nonresponse bias in the P sample. Interviews were conducted during the Person Interview field data collection (refer to Marra and Kennel, 2022, for more information on the Person Interview). Nonresponding households were not interviewed because of one of the following:

- They could not be contacted.
- The interview was refused.
- An interview was conducted but the reported data was not complete enough to uniquely identify anyone in the household.

We implemented the NIA before the P-sample person status imputation and weight trimming. The NIA procedure affected the match rate, the denominator of the dual-system estimates (Marra and Kennel, 2022). The main output of the NIA was a set of noninterview-adjusted weights, the sampling weights multiplied by NIA factors.

The response rate for occupied housing units in the 2020 PES was 83.2 percent for the U.S. For comparison, the 2010 CCM response rate was 96.6 percent for the U.S. Table 3 and Table 4 summarize the results of the Person Interview for the 2020 PES and the 2010 CCM. It should be noted that the fieldwork for the 2020 Person Interview took place during the COVID-19 pandemic. Some sample areas were on lockdown during the initial Person Interview period because of health and safety concerns. The 2020 PES included an additional interview period after the originally planned Person Interview. This additional operation, called the Person Interview Reopen, conducted interviews for noninterviewed households. The aim of the Person Interview Reopen was to increase response rates, though final response rates were still lower than in 2010. All tables include the additional completed interviews after the Person Interview Reopen efforts.

Table 3: Summary of the Person Interview (Unweighted)

Interview Status	2020		2010	
	Count	Percent	Count	Percent
Total Housing Units	161,000	100.0	171,000	100.0
Interview	114,000	70.8	140,000	81.9
Noninterview	23,000	14.3	5,300	3.1
Vacant	19,000	11.8	21,500	12.6
Nonexistent	5,200	3.2	4,700	2.8

Notes:
1. Counts may not sum to totals shown because of rounding.
2. The 2020 PES counts include the Person Interview and the Person Interview Reopen.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (March 2022 release) and 2010 Census Coverage Measurement Survey.

The 2020 PES NIA procedure used a stratified propensity model. Comparisons and explanations of differences in NIA methodology for the 2010 CCM and 2020 PES are described in Jost and Konicki (2018). Although implemented at the housing-unit level, the NIA only affected the P-sample person weights. The NIA factors were applied to P-sample person weights before status imputation and weight trimming for person net coverage error estimation. We did not conduct a noninterview adjustment for housing unit estimation. Housing unit data did not necessarily require an interview and could be obtained via inspection. Missing data for housing units after field operations was handled via imputation rather than weighting adjustment.

2.1 Identifying Noninterviewed Housing Units

Based on the Person Interview field work, each housing unit was assigned an interview outcome:

- Interview.
- Noninterview.
- Vacant.
- Nonexistent (an address determined to be not a housing unit).

Because dual-system estimation had stricter definitions of interviews and noninterviews compared to field operations, some of the field interview outcomes were changed when processing the data for estimation. Vacant and nonexistent units were identified in the field and generally unchanged. They were treated as out-of-scope for the NIA. Interviews and noninterviews were also identified in the field, but the NIA reclassified some of the interviews as noninterviews based on review of the person data[4].

[4] There were also cases where interviews were reclassified as vacant or nonexistent based on the information from the Person Interview. This was typical if it was determined that all the residents were out-of-scope of the PES (vacant) or an address was determined not to be a housing unit (nonexistent).

As described in the Source and Accuracy Statement (Marra and Kennel, 2022), noninterviews were split into two main categories. The first category of noninterviews were cases where no interview was conducted, including cases for which contact was made but there was a refusal or unresolved language barrier. The second category of noninterviews were cases where an interview did take place but there was not enough usable information from the respondent.

An interview was converted to a noninterview in estimation when one of the following occurred:

- All people in the housing unit had insufficient information for dual-system estimation (refer to Section 1.2).
- All people in the housing unit were duplicates[5], fictitious[6], had insufficient information for dual-system estimation, or were a combination of these.
- At least one but not all of the people in the housing unit were out-of-scope and all other people were duplicates, fictitious, or had insufficient information for dual system estimation.

These conditions cover most changes, but a few households changed from an interview to a noninterview, vacant, or nonexistent during estimation based on other information collected during the Person Interview.

Table 4 shows the PES interview status of occupied housing units for the 2020 PES and 2010 CCM. There were more noninterviews in the 2020 PES compared to the 2010 CCM. A major contributor to higher nonresponse in 2020 was a higher number of interviews where all person records within the household had insufficient information for dual-system estimation. Whole households with insufficient person records meant that none of the people in the household were rostered with enough information to perform accurate matching and follow-up. These interviews were not considered sufficient and were converted to noninterviews during estimation. As mentioned previously, interviews could be converted to noninterviews for reasons other than a whole household of insufficient person records. This was much less frequent for both the 2020 PES and 2010 CCM compared to whole households of insufficient information person records.

[5] Duplicates here refer to duplicates of other Person Interview person records.
[6] Fictitious refers to fictional or false person records.

Table 4: Person Interview Response Rates of Occupied Housing Units (Unweighted)

Interview outcome	2020		2010	
	Count	Percent	Count	Percent
Total	137,000	100.0	145,000	100.0
Interview	114,000	83.2	140,000	96.6
Noninterview	23,000	16.8	5,300	3.7
Interview not conducted	4,900	3.6	2,300	1.6
Interview not sufficient	18,500	13.5	3,000	2.1
Whole-household insufficient	16,500	12.0	2,400	1.7
Other	1,700	1.2	600	0.4

Notes:

1. Whole-household insufficient refers to households where all people were rostered with insufficient information for dual-system estimation.
2. The Other category includes all cases where all person records within the housing unit were either fictitious, duplicates, insufficient information, or out-of-scope.
3. Counts may not sum to totals shown because of rounding.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

Table 5 breaks down the whole households with insufficient information into who responded to the interview and whether the interview finished. A proxy respondent was someone outside of the household, a neighbor for example, who answered questions for the sampled housing unit. The completion status of the interview was labeled as either “finished” or “break-off.” Finished means every relevant[7] question was asked during the interview, though not every question was necessarily answered. A break-off interview means only a portion of the questions were asked during the interview. Noteworthy is that the majority of the whole-household insufficient information cases were break-off by household member or break-off by proxy respondent.

[7] During the Person Interview, some questions were asked depending on the demographics of household members. For example, an interviewer would not ask if Duncan, age one month, was attending college.

Table 5: Whole Household of Insufficient Information Person Records by Interview Type (Unweighted)

	2020		2010	
	Count	Percent	Count	Percent
Whole-Household Insufficient	16,500	100.0	2,400	100.0
Household Interview	8,300	50.3	1,400	58.3
Finished	750	4.5	20	0.8
Break-off	7,600	46.1	1,400	58.3
Proxy Interview	8,300	50.3	1,000	41.7
Finished	350	2.1	< 15	D
Break-off	8,000	48.5	1,000	41.7

Notes:

1. Whole-household insufficient refers to households where all people were rostered with insufficient information for dual-system estimation.
2. Finished means that all relevant questions were asked, not all questions were answered.
3. Counts may not sum to totals shown because of rounding.
4. D: Data withheld to avoid disclosure.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

2.2 Overview of the Propensity Modeling Approach

The 2020 PES NIA procedure used a stratified propensity model, instead of the cell-based algorithm used in the 2010 CCM (Jost and Konicki, 2018). The 2010 CCM NIA distributed the weights of occupied P-sample noninterviewed housing units to similar P-sample interviewed housing units within cells. For the 2020 NIA, response was modeled for each housing unit using logistic regression. Housing units were then put into groups, called propensity strata, based on their modeled response propensity. The NIA procedure distributed the weights of noninterviewed P-sample housing units to interviewed P-sample housing units within each propensity stratum. The 2020 PES propensity strata served the same function as the cells in the 2010 CCM in the sense that both were created with the purpose of grouping together similar housing units.

The response propensity was the probability that a household interview was completed and sufficient, given that the housing unit was selected in the sample. The advantage of the propensity models was that they added flexibility to which covariates could be used. Propensity modeling could include or leave out interaction terms, whereas the cell-based method implicitly included all interaction terms. Another benefit was that explicit model fitting techniques could be used to determine the set of covariates in the model. Stratification was chosen since it was less reliant on a correctly specified regression model than directly using the propensities to create the NIA factors (refer to Valliant et al., 2013, for discussion on directly using propensities). This major benefit is why propensity stratification is commonly used for nonresponse adjustments when available model variables tend to be limited. The goal of the NIA was to group housing units with similar characteristics and response propensities together so that the weight of noninterviewed housing units was transferred to similar interviewed housing units.

As in the 2010 CCM, we implemented a separate NIA for American Indian Country[8] (AIC) areas and non-AIC areas. We performed separate adjustments because AIC areas were oversampled (and therefore received smaller weights) compared to non-AIC areas. We also did not want to transfer sampling weights of nonresponding households to responding households from AIC areas to non-AIC areas or vice versa.

2.3 Details of the Noninterview Adjustment

The 2020 PES NIA used a stratified propensity model to distribute sampling weights from noninterviews to interviews. We increased the weights of the interviewed housing units with similar characteristics as the noninterviews to reduce nonresponse bias and increase the representativeness of the sample. We modeled response propensity with a logistic regression model using frame variables. These variables were collected during the initial PES housing unit listing operation and were available for interviewed and noninterviewed housing units.

The logistic regression modeled the final interview status for interviewed and noninterviewed housing units. We fit models separately for non-AIC and AIC areas. Typical of noninterview adjustments, the lack of information on noninterviewed housing units limited the variables available for the model. Many of the same variables used in the 2010 CCM NIA method were chosen as model variables (Konicki et al., 2013). As previously mentioned, stratification alleviated some of the burden on model variables, as the propensities and model variables were indirectly used to create the NIA factors. The propensity model had fixed effects and a random effect. Fixed effects are model variables with set levels. A random effect is a model variable where the levels are randomly selected from a population[9]. The model variables and stratification parameters in the 2020 PES NIA model are listed in Table A1 of Attachment A: Noninterview Adjustment Models and Table D1 of Attachment D: Variable Descriptions.

One important aspect of the models was the inclusion of the block as a random effect. The 2010 CCM equivalent, the block cluster, was used in the 2010 NIA procedure to define cells. We found that the block random effect was statistically significant for modeling propensity. This introduced a “local-geography” influence to the propensities. Therefore, housing units that had

---

[8] American Indian Country includes reservations, off-reservation trust land, tribal statistical areas, and Alaska Native village statistical areas.

[9] For an example of fixed effects and random effects, consider an experiment on high school students’ math grades. The fixed effects could be age and sex and the school could be a random effect. Variables like age and sex have a set of known levels. The schools in the experiment would be sampled from the whole population of schools; the random effect is meant to represent the larger population.

the same fixed effects levels but were in different sampled blocks, received different model propensities. Including the block random effect also added variability to the predicted response propensities[10].

Equation (1) is the form of the logistic regression and shows the contribution of the random effect.

$$logit\big(P(Y=1)\big) = \beta_o + \boldsymbol{\beta X_{ij}^T} + u_j \tag{1}$$

where,

- Y = the interview status (response indicator) for each housing unit.
- $i \in \{1, \dots, I_j\}$ for each housing unit in each random effect level *j.*
- $j \in \{1, \dots, J\}$ for each random effect level.
- $P(Y=1)$ = probability of household interview.
- β_o = fixed intercept parameter.
- $\boldsymbol{\beta}$ = fixed effects parameter vector.
- $\boldsymbol{X_{ij}^T}$ = covariate vector of fixed effects.
- u_j = random effect parameter for each random effect level *j*.

The predicted response propensity (probability of an interview) was calculated by transforming the value calculated in equation (1) for each occupied housing unit as shown in equation (2).

$$PROPENSITY = \frac{exp\big(\beta_o + \boldsymbol{\beta X_{ij}} + u_j\big)}{1 + exp\big(\beta_o + \boldsymbol{\beta X_{ij}} + u_j\big)} \tag{2}$$

The propensity strata were created separately for non-AIC and AIC areas. For the non-AIC areas, the response propensities for all occupied housing units were sorted in ascending order by state and placed into five equal-sized strata of occupied housing units within each state. We used five strata in part based on the variability of the response propensities. This method preserved the total weight of the nation and each state. By creating the adjustment at the state level, people from one state did not represent people in other states and state sampling weight totals were not changed by the adjustment.

The AIC areas were divided into subregions based on states and American Indian Reservations. Propensities were sorted in ascending order within each subregion and then placed in five equal-sized strata within subregion. Since some of the sample American Indian Reservations crossed state lines, the AIC stratification ensured that all American Indian Reservations that crossed state lines were placed in the same subregion.

---

[10] Without the block random effect, there would be few unique propensity values. For example, if the only variables were fixed effects such as housing unit type of address (3 levels) and state (51 levels) then there would be 153 (3*51) total unique values of propensity.

Within each stratum, the sampling weights of the noninterviewed housing units were equally distributed to the unweighted interviewed housing units. To account for the variability of this process, we repeated the adjustment for 80 sets of replicate weights (for more information on replication, refer to Zamora, 2022).

The 2010 CCM NIA included a check in each cell that prevented the weight of many noninterviewed housing units being spread to comparatively few interviewed housing units. In the 2020 PES, we implemented this check after creating initial propensity strata. If the initial stratum had more noninterviews than twice the number of interviews, then the stratum was collapsed with an adjacent stratum. The check was then applied to the new combined stratum. This check was performed iteratively until the number of noninterviews was less than twice the number of interviews in each final stratum.

The last step of the NIA was to create the NIA factor for the full sample and 80 replicates. Within each final stratum, the NIA factor for interviewed housing units was the total occupied sampling weights (including noninterviewed housing units) over the total of the sampling weights for interviewed housing units; refer to Equation (3). The NIA factor for noninterviewed housing units was 0. The vacant and nonexistent housing units were out of scope for the NIA so they did not receive a NIA factor. Every interviewed housing unit in a final stratum received the same NIA factor. The same calculation was performed using the same set of strata for the full sample and the 80 replicates.

$$NIAF = \frac{SUMINTWEIGHTP + SUMNONINTWEIGHTP}{SUMINTWEIGHTP} \tag{3}$$

where,

- SUMINTWEIGHTP is the sum of the weight of interviewed housing units within a given propensity stratum.
- SUMNONINTWEIGHTP is the sum of the weight of the noninterviewed housing units within a given propensity stratum.

2.4 Noninterview Adjustment Results

As previously stated, the NIA weights were used in the P-sample imputation and the match model for dual-system estimation. Table 6 shows a summary of the NIA factors calculated in 2020 and 2010. The mean values of the NIA factors for interviewed housing units were 1.208 and 1.038 for the 2020 PES and 2010 CCM, respectfully. Therefore, the weight of the interviewed housing units increased by 20.8 percent on average in the 2020 PES because of the NIA. In the 2010 CCM, the weight of the interviewed housing units increased by 3.8 percent on average because of the NIA. The larger NIA factors in the 2020 PES were because of a lower response rate than the 2010 CCM.

Table 6: Noninterview Adjustment Factor Distribution for Interviewed Housing Units

Year	Minimum	25th Percentile	Median	Mean	75th Percentile	Maximum
2020	1.000	1.040	1.112	1.208	1.248	4.148
2010	1.000	1.000	1.016	1.038	1.055	3.167

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

3. Logistic Regression Models for Status Imputations for Net Coverage Error

After finishing all data collection activities, there remained E-sample enumerations without enough information to determine the enumeration status, and P-sample people without enough information to determine the inclusion or match statuses. A common reason for unresolved E-sample and P-sample statuses was the lack of reported data from the PES interviews needed to determine the correct enumeration, inclusion, or match status. We imputed values for the missing statuses using survey-weighted logistic regression models fit on the resolved data. This was true for the enumeration status imputations described in Section 4, and also for the inclusion and match status imputations described in Section 5.

Logistic regression modeling was used in the imputation of statuses for the 2010 CCM (Konicki et al., 2013). Logistic regression models are well suited for status imputations. They yield predicted probabilities between zero and one. Equation (1) in Section 2.3 presents the formula for the logistic regression model. The beta terms ($\boldsymbol{\beta}$) represent the covariates. Unlike the NIA propensity models, for the imputation of statuses there was no random effect.

The statuses we imputed were binary, thus all logistic regression models described in this report had binary dependent variables. Resolved cases representing the “yes” category (i.e., included in the P sample, match, or correct enumeration) were assigned a value of 1 for the dependent variable. Resolved cases representing the “no” category (i.e., not included in the P sample, nonmatch, or erroneous enumeration) were assigned a value of 0 for the dependent variable.

The sample design was taken into account when fitting all logistic regression models for imputation by using survey-weighted logistic regression procedures. To account for variability of the imputation, each survey-weighted logistic regression model was fit to the data 81 times – once for the full sample and each other time using a different set of replicate weights.

Missing statuses were imputed with a fraction between 0 and 1. For example, a P-sample case might have a predicted probability of 0.80 for inclusion status, meaning 80 percent of the record’s weight was counted as being in the P sample and 20 percent was not. If we also imputed the match status, some portion of the 80 percent contributed to a match and the reminder was a nonmatch. Thus, 50 percent of the record's weight might be a match, 30 percent a nonmatch, and 20 percent out-of-scope.

The coefficients estimated from the logistic regression models were applied to the corresponding variables of unresolved cases to compute their predicted probabilities of the missing status using Equation (2), without the random effect.

In a final step, we split E-sample records with imputed probabilities of correct enumeration into two records, one a correct enumeration and the second an erroneous enumeration. The correct enumeration record was weighted by the imputed probability of correct enumeration, and the erroneous enumeration record was weighted by the 1 minus the probability of correct enumeration. We did this to facilitate the DSE modeling of probabilities of correct enumeration, which used logistic regression modeling and thus required 0 or 1 for the modeled response data. We did the same for P-sample people with imputed match statuses. Refer to Heim (2022) for details on this process.

3.1 Example of a Covariate: Before Followup Match Code Group

The goal of imputation with logistic regression modeling is to use existing information to make a prediction about the missing statuses. For example, certain demographic covariates have a history of being correlated with both erroneous enumerations and nonmatches. These include owner/renter, age, sex, race, and Hispanic origin. Including these covariates in a logistic regression model yielded better predictions of status than naively substituting an overall mean probability.

A covariate that, by itself, had noticeable predictive power was the Before Followup Match Code Group (BFUMCG). This variable existed in different forms for both the E sample and the P sample, though we only discuss the E-sample variable here. Table 7 shows the distribution of correct and erroneous enumerations by the values of BFUMCG. During clerical matching every census record was reviewed, and staff determined if a follow-up interview was needed to get more information about the person's enumeration status. The Before Followup Match Code Group summarized why a follow-up interview was or was not necessary. We see that Resolved Before Followup had a higher rate of correct enumeration than some other values such as Conflicting Household, Partial Household Nonmatch, and Whole Household Nonmatch.

One can better understand the predictive power of the covariate BFUMCG with some knowledge of the PES processing. E-sample enumerations that matched to a valid and nonmover P-sample person were coded as a correct enumeration and not sent to follow-up (they were BFUMCG Resolved Before Followup) because we already received an independent verification of the enumeration from the Person Interview field operation. That is, the E-sample match to a valid P-sample person indicated that the E-sample enumeration also represented a valid person or correct enumeration. Erroneous enumerations, on the other hand, often could not be matched to valid P-sample people. Thus the E-sample enumerations that did not match to a valid P-sample person had higher probabilities of being erroneous. These nonmatching E-sample cases included the BFUMCG groups Conflicting Household, Partial Household Nonmatch, and Whole Household Nonmatch.

Note that most of the approximately 21,500 E-sample enumerations in the Before Followup Match Code Group labeled Unclassified Inclusion Status of Matching P-sample initially had an enumeration status assigned, but were blanked out in response to concerns about their initial status assignment. These enumerations were matched to P-sample person listings for which we had insufficient information. The PES should have sent them to a follow-up interview to determine the correct enumeration status of the E-sample enumeration, but failed to do so. Since we were uncertain that the P-sample person was a valid person, we could not assume a matching E-sample enumeration was a correct enumeration. So we imputed their enumeration status.

The Before Followup Match Code Group effectively partitioned the resolved cases into cells with different correct enumeration rates. For example, the category Resolved Before Followup had a very high correct enumeration rate, 97.4 percent, while the partial and whole household nonmatch groups had lower correct enumeration rates, 88.2 percent and 88.9 percent, respectively. This kind of partitioning of the data into groups with similar correct enumeration rates within the group, but with differing rates between the groups, is a key characteristic of a powerful covariate for imputation.

To assess the usefulness of a covariate it is also important to consider the distribution of the unresolved cases. Categories that have high numbers of resolved cases but few unresolved cases can yield a logistic regression model with a high-level of fit, and yet be of minimal predictive value in practice. We see this with the two groups with the largest number of resolved cases, Resolved Before Followup and Insufficient Information. They had only 750 and 0 unresolved cases each, respectively. However, some of the other groups demonstrate the potential utility of Before Followup Match Code Group. For example, there were about 15,000 unresolved cases with Whole Household Nonmatch, with a resolved correct enumeration rate of 88.9 percent, and about 21,500 unresolved cases with Unclassified Inclusion Status of Matching P-sample People, with a resolved correct enumeration rate of 97.1 percent.

Table 7: Counts of Correct and Erroneous Enumerations by Before Followup Match Code Group

Before Followup Match Code Group	Correct Enumerations	Erroneous Enumerations	Resolved Correct Enumeration (Percent)	Number of Unresolved
Resolved Before Followup	229,000	6,100	97.4	750
Possible Matches	850	30	96.6	150
Conflicting Household	4,800	500	90.6	3,000
Partial Household Nonmatch	16,500	2,200	88.2	4,800
Whole Household Nonmatch	33,000	4,100	88.9	15,000
Duplicate	2,500	650	79.4	700
Unclassified Inclusion Status of Matching P-sample	10,000	300	97.1	21,500
Insufficient Information	0	40,000	0.0	0
Total	297,000	54,000	84.6	46,000

Note: Counts may not sum to totals shown because of rounding.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

4. Imputation of E-Sample Correct Enumeration Status for Net Coverage Error

To calculate the dual-system estimates, we needed to assign an enumeration status to each E-sample person enumeration. We defined an E-sample enumeration as a **correct enumeration** for the estimation of net coverage error if it was an enumeration with sufficient information that corresponded to a person who should have been counted in the block search area in a housing unit on Census Day. We use this definition of correct enumeration for the rest of this section; in the section on components we have a slightly different definition of correct enumeration.

Enumerations not meeting these criteria were **erroneous enumerations**. Erroneous enumerations included people who were born after Census Day or who died before Census Day, fictitious enumerations, and people counted in the wrong location (i.e., people who should have been counted somewhere outside of the block search area). In addition, if two enumerations referred to the same person, one was called correct and the other erroneous because of duplication.

Note that E-sample enumerations with insufficient information were treated as erroneous enumerations for net coverage error estimation (though not for components of coverage; refer to Section 6). We did this because we could not match P-sample people to them accurately. Some of the P-sample records who represented the same person as an insufficient information census enumeration could be matched. But other P-sample records who represented the same person as an insufficient information census enumeration would not be matched and would yield false nonmatches. False nonmatches would bias the match rate and the DSE used to calculate the census net coverage error. We avoided introducing this bias in the DSE by treating all E-sample insufficient information cases as erroneous enumerations and all P-sample people who matched to insufficient information census enumerations as nonmatches.

Note that whole-person census imputations were not in the E sample.

Table 8 has a summary of the enumeration status for E-sample people. The rate of missing status in the E sample was higher for the 2020 PES than the 2010 CCM. Part of the increased rate resulted from the special fix of E-sample people that matched to P-sample people with insufficient information person statuses who did not go to follow-up (as discussed in Section 3.1). However, even without this fix the E-sample unresolved rate would have been noticeably higher than the 2010 CCM rate. There are several reasons that may have contributed to the higher unresolved rates in the 2020 PES. They include difficulty conducting interviews because of COVID-19, the greater amount of missing characteristics of E-sample enumerations that made matching and follow-up more difficult, and more insufficient information cases in the P sample, which would have required more E-sample enumerations going to a follow-up interview.

Table 8: 2020 PES and 2010 CCM Person Enumeration Status

	2020 PES	2010 CCM
Total E-sample Enumerations	397,000	384,000
Number of Resolved Enumerations	351,000	365,000
Number of Unresolved Enumerations	46,000	18,500
Unresolved Enumeration (Percent)	11.6	4.8

Note: Counts may not sum to totals shown because of rounding.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

4.1 Logistic Regression Modeling to Impute for Missing Correct Enumeration Status

We imputed the probability of correct enumeration using logistic regression for cases with a missing enumeration status. We used one of two models to predict the correct enumeration probability for unresolved E-sample enumerations. The models were fit on the resolved cases (there were 351,000 resolved cases in the 2020 PES; refer to Table 8). Each used the same core set of main effects, which included demographic characteristics such as age and sex as well as a proxy interview flag. A full listing of the variables used in each model can be found in Attachment B: E-Sample Imputation Models. A description of each variable and its categories can be found in Table D1 in Attachment D: Variable Descriptions. The weight used in the models was the original E-sample weight (the inverse of the probability of selection).

The first model included nine additional indicator variables, while the second model excluded these variables. One of these variables was a duplicate link flag that indicated whether the E-sample person was linked to another census enumeration as a possible duplicate. The other eight flag variables indicated whether a person had certain types of additional addresses attached to them. These additional address flags included outmover, seasonal, inmover, college, relative, military, job, and group quarters address flags.

We made this distinction with the existence of an address flag because we had a clear response for a "yes" response. In contrast, if a person record did not have any additional addresses attached to them, it was not clear if the respondent did not have an alternative address or was not responding to the question. The respondent might have neglected to provide this additional information or might not have known this information for the people they were responding for. We did not want the predicted probability for people without the address flags to be influenced by the effects of the address flags in the model.

4.2 Final Correct Enumeration Probability

After the imputation process, each E-sample person enumeration was assigned a temporary correct enumeration probability, P_{tce}.

$$P_{tce} = \begin{cases} \mathit{1, if\ person\ was\ correctly\ enumerated} \\ \mathit{0, if\ person\ was\ erroneously\ enumerated} \\ P^{*}_{tce}, \mathit{if\ enumeration\ status\ was\ unresolved} \end{cases} \tag{4}$$

For the unresolved cases, P^{*}_{tce} was assigned from one of the two logistic regression models as previously described.

We made an adjustment to the correct enumeration probability of an E-sample enumeration based on the count of duplicates with enumerations subsampled out of the E sample in large blocks (blocks with 58 or more housing units). If there was a duplicate between an E-sample person and one or more subsampled out enumerations, as a rule we assigned the E-sample enumeration a code of correct enumeration and a temporary probability of correct enumeration of 1. The subsampled out enumeration(s) received the code of erroneous enumeration because of duplication. However, this rule would always yield a probability of correct enumeration of 1 for the E-sample enumeration, when in expectation half the time it should be erroneous because of duplication. Thus, to avoid introducing bias, we multiplied the temporary correct enumeration probability by an adjustment factor as shown in equation (5) to get the final probability. Note that this adjustment also took into account situations where there were three or more enumerations referring to the same person.

$$P_{fce} = P_{tce} \times \left(\frac{n_{es} + 1}{n_{es} + n_{nes} + 1}\right) \quad (5)$$

where,

- P_{fce} is the final correct enumeration probability.
- P_{tce} is the temporary correct enumeration probability.
- n_{es} is the number of duplicate links of the E-sample enumerations to other E-sample enumerations in the block.
- n_{nes} is the number of duplicate links of the E-sample enumerations to census enumerations not in the E sample but in the block.

4.3 Results for Imputing the Correct Enumeration Status

Table 9 shows the overall effect of imputation on the correct enumeration rate. While the difference could appear small at first glance, at the national level it would have a noticeable impact. Note that the overall correct enumeration rate presented in Table 9 differs from estimates of the component "correctly enumerated in the BCU search area," as presented in Table 2 of Khubba et al. (2022), or in later tables in this report, such as Table 20. The component of coverage estimates included imputations for E-sample insufficient information cases, whereas the estimates of net coverage error treated the E-sample insufficient information as erroneous. Imputation increased the correct enumeration rate from 86.75 percent to 87.16 percent. This increase in the correct enumeration rate resulted in an increase in the dual-system estimate of the population size.

Table 9: Correct Enumeration Rate With and Without Imputation for the 2020 PES (Weighted)

	Resolved Cases Only	After Imputation
Correct Enumeration Rate in Percent	86.75	87.16

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

Of course, the imputed correct enumeration rates varied. Table 10 shows the 25th percentile, the median, and the 75th percentile of the imputed values.

Table 10: Distribution of Imputed Correct Enumeration Probabilities for the 2020 PES (Unweighted)

25th Percentile	Median	75th Percentile
0.8761	0.9427	0.9785

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

5. Imputation of P-Sample Inclusion, Mover, and Match Statuses for Net Coverage Error

Dual-system estimation required us to determine whether each person in the P sample matched to an enumeration in the Census[11]. After all PES data collection activities were completed, there remained people listed in the P sample without enough information to determine an inclusion, mover, or match status. This section provides an overview of issues pertaining to missing statuses in the P sample.

5.1 The P-Sample Statuses

There were four P-sample statuses relevant to dual-system estimation: inclusion, mover, inmover match, and nonmover match. The first we discuss is the inclusion status. Before determining the match status, we needed to determine which PES person listings were in-scope for the P sample. People living in Group Quarters facilities (for example a prison, college dorm, or nursing home) and Remote Alaska areas on Census Day, and visitors are examples of people who were not eligible to be in the P sample. Sometimes we did not have enough information to determine if someone should have been included in the P sample. Thus, we imputed the inclusion status for such people.

The PES Person Interview included questions about where everyone in the household on the PES Interview day was living on Census Day. The responses to these questions were reviewed by clerical technicians to identify if each person in the household on the PES Interview day was in-scope for the PES. If we did not have enough information to determine if the person was in-scope for the PES, we imputed the P-sample inclusion status.

Once the inclusion status was determined for all people listed in the PES Person Interview, we had to determine their mover status. It was possible that the person moved into the PES housing unit between Census Day and the PES Interview Day, in which case they were an **inmover**. If the person lived at the same address on both Census Day and Interview Day, they were a **nonmover**. The PES Person Interview asked questions about where people were living around Census Day. The reported Census Day addresses were compared to the Interview Day address by clerical matching staff. If the clerical matching staff were not able to determine if a person was an inmover or nonmover, we imputed their mover status.

Mover status was important because it determined the search area for matching the PES to the Census (refer to the next paragraph for details). In past post-enumeration surveys, mobility has been a major factor in our ability to determine match status. Inmovers were generally more difficult to match to the census than nonmovers and had higher unresolved match rates. For this reason, we imputed match status separately for inmovers and nonmovers. And, because

[11] Refer to Zamora (2022) for details on how the dual-system estimates were calculated.

we did not always know whether a person was an inmover or nonmover, we had to impute mover status before imputing match status. Thus, P-sample people required up to four separate imputations:

- Inclusion status: Whether a person listed during the PES Person Interview should have been included in the P sample.
- Mover status: Whether a person was a nonmover—i.e., lived at the PES sample address on both Interview Day and Census Day—or whether they were an inmover—i.e., moved into the PES address after Census Day.
- Inmover match status: Given that a person was an inmover, whether they matched to a census enumeration at their Census Day address.
- Nonmover match status: Given that a person was a nonmover, whether they matched to a census enumeration at the PES sample address.

To limit the error of false matches (calling a P-sample person and census enumeration a match when they referred to different people), the matching was done in a limited search area. The search area for an inmover consisted of the block containing the address they reported being at on Census Day and the ring of surrounding blocks. The search area for a nonmover consisted of the block containing the PES sample address and the ring of surrounding blocks. A P-sample person was considered a match only if they matched to a census enumeration in the correct search area. If they matched to an enumeration outside the search area, they were classified as a nonmatch for dual-system estimation. For more information on the PES search area, refer to Hogan (2003).

The following equation indicates how information on the statuses was combined to calculate the overall probability that a P-sample person matched to a census enumeration.

$$p_{match,j} = p_{inmover,j} \times p_{match|inmover,j} + \left(1 - p_{inmover,j}\right) \times p_{match|nonmover,j} \qquad (6)$$

where for person record j,

- $p_{match,j}$ is the overall probability of being a match.
- $p_{inmover,j}$ is the probability of being an inmover.
- $p_{inmover\ match,j}$ is the inmover match probability.
- $p_{nonmover\ match,j}$ is the nonmover match probability.

5.2 Unresolved P-Sample Statuses

A person for whom it was possible to determine a given status is referred to as "resolved" for that status. For instance, people with a resolved inclusion status were those that were identified either as in the P sample or as not in the P sample. It was not always possible to determine the inclusion, mover, inmover match, or nonmover match status of a person listed during the Person Interview—rendering them "unresolved" for that status or those statuses. It was possible that a P-sample person could be resolved for one or more status but not for others.

P-sample people with at least one unresolved status fell into one of two categories:

- **Sufficient information** for dual-system estimation.
- **Insufficient information** for dual-system estimation.

Refer to Section 1.2 for the definitions of sufficient and insufficient information for dual-system estimation.

Table 11 presents counts and rates of P-sample cases that were missing each status. The first row, "P-Sample Inclusion Status," shows the raw counts. However, the counts for mover status and the match statuses were multiplied by the person's probability of being in the P sample (imputed to be greater than 0 but less than 1 for those with an unresolved inclusion status). For example, consider how we obtained the mover status counts. Of the records with a resolved inclusion status, around 262,000 were known to be in the P sample (refer to Table 14) and thus counted as one record each in the mover status calculations. These records were summed with the roughly 21,000 records with an unresolved inclusion status, multiplied by their respective predicted inclusion probabilities. The predicted inclusion probabilities equaled approximately 0.80 on average, leading to the Mover Status total of 279,000. The match status counts additionally account for the probability of being an inmover (imputed between 0 and 1 for those with an unresolved mover status).

Table C1 in Appendix C shows the raw counts for each imputation step (i.e., counts not modified based on inclusion or mover probabilities). In this table, each record that was unresolved for a given status is counted as full record when calculating values for a subsequent status. For example, consider a person with an unknown inclusion and mover status. If we imputed an inclusion probability of 0.70 for the person, the 'Total' value in the mover status row of Table 11 would increase by 0.70, but the 'Total' value in the mover status row of Table C2 would increase by 1. We present both tables because the raw counts in Table C3 and the counts modified for differences in inclusion and inmover probabilities in Table 11 illustrate different aspects of the imputation.

The results show that the PES had to rely on imputation procedures to a greater degree than the 2010 CCM because of higher rates of people with unresolved statuses. Indeed, the overall unresolved inclusion rate was over twice as large in 2020 as 2010 (6.98 percent vs. 2.87 percent) as was the unresolved match rate (5.02 percent vs. 1.90 percent)[12].

[12] These percentages account for differences in estimated inclusion and mover probabilities. Refer to Table C1 in the Attachment for the raw numbers of cases with unresolved statuses.

Table 11: Unresolved Rates for Inclusion, Mover, and Match Statuses (Unweighted)

	2020				2010			
	Total	Resolved	Unresolved	Unresolved (Percent)	Total	Resolved	Unresolved	Unresolved (Percent)
P-Sample Inclusion Status	301,000	279,000	21,000	6.98	383,000	372,000	11,000	2.87
P-Sample Mover Status	279,000	264,000	15,500	5.56	352,000	345,000	7,600	2.16
P-Sample Total Match Status	279,000	265,000	14,000	5.02	352,000	346,000	6,700	1.90
P-Sample Inmover Match Status	22,000	16,500	5,400	24.55	28,000	24,500	3,300	11.79
P-Sample Nonmover Match Status	257,000	249,000	8,600	3.35	324,000	321,000	3,300	1.02

Notes:

1. Counts for the mover status row are multiplied by the probability of inclusion.
2. Counts for the inmover match status row are multiplied by the probability of inclusion and the probability of being an inmover.
3. Counts for the nonmover match status row are multiplied by the probability of inclusion and the probability of being a nonmover.
4. Counts for total match status were calculated by summing the respective counts in the inmover match status and nonmover match status rows. This approach is valid because the inmover match status and nonmover match status figures are multiplied by, respectively, the inmover and nonmover probabilities.
5. Counts may not sum to totals shown because of rounding.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

5.3 Logistic Regression Modeling to Impute for Missing P-Sample Statuses

As with the imputation for correct enumeration status (Section 4), we used survey-weighted logistic regression models to impute missing P-sample statuses. For fitting the inclusion status models, the weight used was the product of the sampling weight and the noninterview adjustment factor. For the mover and match models, the weight used was the product of the sampling weight, the noninterview adjustment factor, and the final inclusion probability[13].

For each status, different models were used to impute sufficient and insufficient information cases. The chief distinction between sufficient and insufficient information models was that the sufficient information model relied on the P-Sample Before Followup Match Code Group. The Before Followup Match Code Group distinguished cases based on whether they went to Nonresponse Followup, classifying those that did not go according to their resolved designation (e.g., nonmover, inmover, out-of-scope) and classifying those that did go to follow-up based on their reason for going (e.g., possible match, conflicting household). P-Sample Before Followup Match Code Group was the most predictive covariate in our sufficient information models. However, insufficient information cases were not sent to follow-up, therefore we could not use this predictor in the respective models. Instead, some form of the insufficient information match code group variable—based partly on housing unit match status—was used in each of the insufficient information models. A full listing of the variables used in each model can be found in Table C2 of Attachment C: P-sample Imputation Models. A description of each variable and its categories can be found in Table D1 in Attachment D: Variable Descriptions[14].

Table 12 presents the weighted match rate before and after the status imputation processes. The "Resolved Match Rate" is limited to people for whom inclusion status, mover status, and match status could be directly measured because they were not missing any data required for determining these statuses. The "After Imputation" is the match rate including unresolved cases after all imputation was finished. As we see, the imputation decreased the match rate from 86.77 percent to 84.98 percent. This decrease in the match rate increased the dual-system estimate of the population size.

---

[13] Note that this logistic regression-based approach to mover status imputation departed from the procedure used in the 2010 CCM. In 2010, a cell mean methodology was used to impute mover status probabilities for unresolved cases, where the cells were based on the BFU match group variable for sufficient information cases and on the BFU insufficient information group variable for insufficient information cases.

[14] In addition to the main set of results based on the full sample, we calculated 80 sets of replicate results following the same procedure. Replicate values were used to obtain standard errors for PES estimates that account for the survey design. Refer to Zamora (2022) for details.

Table 12: Match Rate With and Without Imputation for the 2020 PES (Weighted)

	Resolved Cases Only	After Imputation
Match Rate (Percent)	86.77	84.98

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

Table 13 presents the median and the 25th and 75th percentiles of the imputed probabilities for a P-sample match status. The median imputed probability was 0.7195. One quarter of imputed match probabilities were below 0.3251 while another quarter were above 0.8542.

Table 13: Distribution of Imputed Match Probabilities for the 2020 PES

25th Percentile	Median	75th Percentile
0.3251	0.7195	0.8542

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

5.3.1 Inclusion Status Imputation

For records with a missing P-sample inclusion status, we imputed an inclusion probability using one of three survey-weighted logistic regression models. Two of the three models were used to impute inclusion probabilities for unresolved sufficient information cases and the third was used to impute inclusion probabilities for insufficient information cases[15]. Each model was fit on the 279,000 cases with a resolved inclusion status (refer to Table 11), i.e., people determined either to be in the P sample or determined to be out-of-scope for the P sample.

The first and second sufficient information models were quite similar except that the first included three address flag variables indicating if a person had certain types of additional addresses attached to them. During the Person Interview, respondents could report addresses other than the sample address with which they were associated. It seemed likely that having such an alternate address could be associated with lower probability of being in scope for the P Sample. To account for this potential relationship, we included in the first sufficient information model mover and seasonal address flags, as well as an “other address” flag that was a combination of other address types such as college, military, and jail. The second sufficient information model omitted these alternate address variables. If an unresolved sufficient information case had at least one of the three flags, it received the predicted

[15] A note about people in noninterviewed, vacant, or deleted housing units: they were neither used to estimate the inclusion status imputation models nor had their inclusion statuses imputed by these models. These three types of records had their final inclusion probability set to 0.

probability from the first sufficient model. Otherwise, the unresolved sufficient information case received the predicted probability from the second model. We had separate models for people with and without address flags for the same reasons we did when modeling the correct enumeration status (refer to Section 4.1).

The third model was used to impute inclusion probabilities for insufficient information cases. This model, though using many of the same variables as the sufficient information models, did not include the three address flags and relied on the BFU insufficient information group variable instead of the BFU match code group. As was previously discussed, while the Before Followup Match Code Group variable had a very strong association with inclusion status and other outcomes among resolved cases, it could not be used to predict outcomes for insufficient information cases because these cases were not sent to follow-up. Refer to Table C2 in Attachment C: P-sample Imputation Models for a list of the predictor variables used in each of the inclusion status models and Table D1 in Attachment D: Variable Descriptions for a description of these covariates.

Table 14 provides a summary of the results of the inclusion status imputation. The first row shows cases with a resolved inclusion status; 93.9 percent were classified as being in the P sample. By comparison, the missing inclusion status cases were imputed with a lower inclusion rate of 81.0 percent overall. When we break down these cases by sufficiency status, we imputed a P-sample inclusion rate of 85.1 percent for sufficient information cases, and 75.2 percent for insufficient information cases. We would not expect the people missing the inclusion status to have as high a rate of inclusion as the resolved cases because many of the people with an unresolved inclusion status had characteristics that were similar to cases that were classified as not in the P sample after follow-up interviews.

Table 14: Inclusion Rates of Resolved and Unresolved P-Sample Cases (Unweighted)

	Total	Included	Not Included	Inclusion (Percent)
Resolved Inclusion Status	279,000	262,000	17,000	93.9
Unresolved Inclusion Status	21,000	17,000	4,300	81.0
Sufficient Information	8,700	7,400	1,300	85.1
Insufficient Information	12,500	9,400	3,100	75.2

Notes:

1. For people with an unresolved P-sample inclusion status, counts of included cases were estimated as the sum of these individuals' imputed probabilities of being included in the P sample.
2. Counts may not sum to totals shown because of rounding.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

5.3.2 Mover Status Imputation

For P-sample people missing the mover status, we imputed a mover probability using one of two survey-weighted logistic regression models. The first model was used to impute mover probabilities for unresolved sufficient information cases, while the second was used for insufficient information cases. Both models were fit using the 264,000 people with a resolved mover status, that is, categorized as being either an inmover or a nonmover (refer to Table C1).

Some people with an unresolved inclusion status had a resolved mover status and thus could be used in the mover status model estimation. Specifically, people whose PES Interview Day residence could not be determined were assigned an unresolved inclusion status. Nevertheless, they were also classified as nonmovers if they were known to have lived in the sample search area on Census Day, and as inmovers if they were known not to have lived there on Census Day. Such cases were accounted for in the mover models by adjusting their weights according to their imputed inclusion probabilities.

Unlike the corresponding inclusion models, the sufficient information mover model relied on the binary Followup Flag variable rather than the more detailed BFU Match Code Group. This decision was based on a concern that – within BFU match code group categories – cases that were resolved through follow-up may have had a lower inmover rate on average than those which remained unresolved. In this situation, imputation using the BFU match code group variable could lead to under-estimation of inmover probabilities. To limit potential inaccuracy of this type, the Followup Flag variable was used instead. Be that as it may, the insufficient information mover model used the BFU Insufficient Information Group variable like the previously-discussed insufficient information inclusion model. Refer to Table C2 in Attachment C: P sample Imputation Models for a list of the predictor variables used in each of the mover status models and Table D1 in Attachment D: Variable Descriptions for a description of these covariates.

The mover status imputation assigned a predicted probability of being an inmover to each P-sample person who had both a nonzero inclusion probability and missing mover status. The model depended on whether the person record had sufficient or insufficient information. Table 15 provides the results of the final mover status assignment. While 7.20 percent of cases with a resolved mover status were determined to be movers, an appreciably larger share of unresolved cases – 18.71 percent – were imputed to be movers. The differences among unresolved cases were even larger, with the inmover rate of sufficient information cases imputed at 33.87 percent compared to 9.04 percent for insufficient information cases. A key aspect of the relatively high imputed inmover rate for unresolved sufficient information cases was that these cases were more likely than resolved sufficient information cases to go to follow-up, and going to follow-up was positively associated with the probability of being an inmover in the sufficient information model. Differences between insufficient information cases and resolved cases in the insufficient information group variable also tended to increase the imputed inmover rate for insufficient cases, but not to the same extent.

Table 15: Mover Rates of Resolved and Unresolved P-Sample Cases (Unweighted)

	Total	Inmover	Nonmover	Inmover (Percent)
Resolved Mover Status	264,000	19,000	245,000	7.20
Unresolved Mover Status	15,500	2,900	12,500	18.71
Sufficient Information	6,200	2,100	4,100	33.87
Insufficient Information	9,400	850	8,600	9.04

Notes:

1. Resolved and unresolved counts are multiplied by the probability of inclusion. For people with an unresolved mover status, counts of movers were estimated as the sum of these individuals' imputed probabilities of being an inmover.
2. Counts may not sum to totals shown because of rounding.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

5.3.3 Inmover Match Status Imputation

For P-sample people missing an inmover match status, we imputed an inmover match probability using one of three survey-weighted logistic regression models. Two of the three models were used to impute match probabilities of unresolved sufficient information cases, while the third was used to impute match probabilities for insufficient information cases. All three models were fit using the 16,500 people in the P sample classified as a matching inmover or as a non-matching inmover (refer to Table C1). Some of these cases had an imputed inclusion status yet also a resolved mover status (refer to preceding section for a discussion of this case type) that allowed their inmover match status to be determined. In estimating the inmover match models, the weight of such cases was adjusted by their imputed inclusion probability. Nevertheless, no case that had an unresolved mover status was assigned a resolved inmover match status (because of the challenge of identifying a search area) and so this subset of records had no influence on the modeling results.

Modeling decisions for inmover match status were shaped by the limited number of inmovers available in the data. In contrast to how inclusion status was handled, the BFU Match Code Group variable was not used to impute inmover match status for sufficient information cases because there were some levels which had too few resolved cases to estimate model coefficients. Instead, the first sufficient information inmover match status model used a partially collapsed BFU match code group variable while the second model instead relied on the Followup Flag variable. The collapsed BFU match code group alleviated many of the estimation problems relating to small sample sizes. However, even this variable had small numbers of resolved inmover match status cases for some of its levels, motivating the use of the even more general Followup Flag as a supplemental variable. Finally, the third model – predicting the inmover match probability for insufficient information cases – used a partially collapsed version of the insufficient information groups variable to contend with categories containing an inadequate number of resolved cases. Refer to Table C2 in Attachment C: P-sample Imputation

Models for a list of the predictor variables used in each of the inmover match status models and Table D1 in Attachment D: Variable Descriptions for a description of these covariates.

Table 16 provides the results of the final inmover match status assignment. In terms of cases with a resolved inmover match status, 72.73 percent were found to match to a census enumeration in the respective search area. The inmover match rate was imputed to be lower for unresolved cases - 64.81 percent. Unresolved sufficient information cases had a similar imputed match rate as insufficient cases (64.44 percent vs. 64.71 percent). A factor which contributed to the lower match rates of unresolved sufficient and insufficient cases was that the Person Interviews associated with them were less likely than those of resolved cases to have been conducted with a household respondent, which is a circumstance positively associated with an inmover match in all three models.

Table 16: Match Rates Conditional on Being an Inmover for Resolved and Unresolved P-Sample People (Unweighted)

Match Status for Inmovers	Total	Match	Nonmatch	Match Rate (Percent)
Resolved Match	16,500	12,000	4,800	72.73
Unresolved Match	5,400	3,500	1,900	64.81
Sufficient Information	4,500	2,900	1,600	64.44
Insufficient Information	850	550	350	64.71

Notes:

1. Resolved and unresolved counts are multiplied by the probability of inclusion and the probability of being an inmover. For people with an unresolved inmover match status, match counts were estimated as the sum of these individuals' imputed probabilities of being an inmover match.
2. Counts may not sum to totals shown because of rounding.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

5.3.4 Nonmover Match Status Imputation

For P-sample people with an unresolved nonmover match status, we imputed a match probability using one of two survey-weighted logistic regression models. The first model imputed match probabilities of unresolved sufficient information cases, while the second imputed match probabilities for insufficient information cases. Both models were fit to people in the P sample classified as a matching nonmover or as a non-matching nonmover – 252,000 in all (refer to Table C1).

Like the inmover match models, the nonmover match models included people with an unresolved inclusion status, with the weights of such cases adjusted to reflect their imputed inclusion probability. Unlike the inmover match models, though, the nonmover match models also included people with an unresolved mover status. The reason for this was that a given P-sample person's nonmover match search area was just the sample block search area. We searched this area for a matching census enumeration for every sufficient information record that *potentially* belonged to a nonmover – regardless of whether the person's Census Day address or Interview Day address had been resolved. A consequence was that there were few (<100) sufficient information cases that had an unresolved nonmover match status.

The sufficient information nonmover match status model used a collapsed version of the BFU Match Code Group variable to avoid estimation problems resulting from some of the original match group categories having small numbers of resolved cases. Likewise, the insufficient information model used a collapsed version of the BFU Insufficient Information Group variable to avoid small cell sizes. The sufficient and insufficient models were quite similar apart from these covariates. For a list of the predictor variables used in each of the nonmover match status models, refer to Attachment C: P-sample Imputation Models, and refer to Table D1 in Attachment D: Variable Descriptions for a description of these covariates.

Table 17 provides the results of the final nonmover match status assignment. Among cases with a resolved nonmover match status, 85.54 percent were found to be a match to a census enumeration. The imputation process resulted in unresolved cases having a lower overall match rate of 79.07 percent. This figure is close to the match rate for insufficient information cases (77.91 percent) because there were few sufficient information cases with an unresolved nonmover match status.

Table 17: Match Rates Conditional on Being a Nonmover for Resolved and Unresolved P-Sample Cases (Unweighted)

Match Status for Nonmovers	Total	Match	Nonmatch	Match (Percent)
Resolved Match	249,000	213,000	36,000	85.54
Unresolved Match	8,600	6,800	1,900	79.07
Sufficient Information	70	30	40	42.86
Insufficient Information	8,600	6,700	1,900	77.91

Notes:

1. Resolved and unresolved counts are multiplied by the probability of inclusion and the probability of being a nonmover. For people with an unresolved nonmover match status, match counts were estimated as the sum of these individuals' imputed probabilities of being a nonmover match.
2. Counts may not sum to totals shown because of rounding.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

6. Enumeration Status Imputation for Components of Census Coverage

In addition to estimates of net coverage error, the 2020 PES produced estimates of components of census coverage for people. This section describes the methodology for imputing missing data to support the estimates of components of census coverage. A key point is that the estimates of the components of coverage were obtained from eight enumeration statuses. It was these statuses that could be missing and for which we had to impute missing data. We used cell means methods for this imputation instead of the logistic regression modeling we used in the imputations for net coverage error, because the multiple response categories did not lend itself to logistic regression modeling.

There was an important difference between the way we handled missing data for the estimation of components from that of net census coverage. For the estimation of net coverage error we treated all E-sample enumerations with insufficient information as erroneous enumerations. In contrast, for the estimation of components of coverage, we assigned a component to person enumerations with insufficient information when we could do so. When we could not assign a component to an E-sample enumeration with insufficient information, we used imputation methods described in this section.

6.1 Components of Census Coverage

The components of census coverage are shown in Table 18. The 2020 PES estimates of these components are found in Khubba et al. (2022).

Table 18: Components of Census Coverage

Correct enumerations	
	Enumerated in the same block[1]
	Enumerated in the same county, though in a different block
	Enumerated in the same state, though in a different county
	Enumerated in a different state
Erroneous enumerations	
	Because of duplication
	For reasons other than duplication
Whole-person census imputations	

[1]More precisely, enumerated in the search area for the correct basic collection unit.

These components detailed why an enumeration was erroneous, or, if it was correct, where the person was counted relative to where they should have been counted. For components of coverage we used a different definition of correct enumeration than that used for net coverage error. For the estimation of components of coverage, any E-sample person who should have been counted in the 2020 Census in a housing unit was considered a correct enumeration. If a person was enumerated outside of the block search area, the components of coverage detailed where the person should have been counted. For example, a person may have been correctly enumerated in the wrong county but in the correct state. Further, if we determined that two enumerations referred to the same person, one was counted as correctly enumerated and the other as an Erroneous Enumeration because of Duplication. Enumerations determined to be Erroneous for Reasons Other than Duplication included situations such as people who were born after Census Day (April 1, 2020), died before Census Day, and fictitious person enumerations.

Note that whole-person imputations were not part of the E sample and the count of whole-person census imputations was obtained from the census.

6.2 Enumeration Statuses for Estimating Components of Census Coverage

The six components of correct enumerations and erroneous enumerations were estimated using the E sample only, based on eight mutually exclusive and exhaustive enumeration statuses. To support the estimates of components of census coverage, the PES collected geographically-detailed information as to where the E-sample person should have been counted on Census Day. This information created these eight enumeration statuses, which are listed in Table 19. For resolved cases, each enumeration status was a “yes” for only one enumeration status and a “no” for the other seven.

Four components of census coverage were directly estimated by an enumeration status: Correctly Enumerated in the Block Search Area (1), Correctly Enumerated in the Wrong State (6), Erroneously Enumerated because of Duplication (7), and Erroneously Enumerated because of Reasons Other than Duplication (8). However, to obtain the estimate of the component Correctly Enumerated in the Correct County but Outside of the Block Search Area, enumeration statuses (2) and (3) were added together in the tabulation. And to obtain the estimate of the component Correctly Enumerated in the Correct State but in the Wrong County, enumeration statuses (4) and (5) were added together in the tabulation.

These eight enumeration statuses were originally designed to produce estimates of an additional component of coverage, Correctly Enumerated in the Correct Place. However, for the 2020 PES we did not produce estimates of this component. (A place is a census geographical concept. Typical places include cities, towns, and communities; refer to U.S. Census Bureau, 1994).

6.3 Missing Data Results for Enumeration Statuses for Components of Census Coverage

When PES operations could not determine some or all of an E-sample enumeration's statuses, those statuses were left missing and said to be unresolved. For such enumerations each enumeration status was either a "no" or missing. That is, some enumeration statuses could be known while others were unresolved. Fully resolved enumerations had one "yes" status and seven "no" statuses.

Table 19 presents the unweighted percentages of unresolved enumeration statuses. The first six enumeration statuses classified an enumeration as correct at the national level; that is, these people were correctly enumerated in a housing unit on Census Day, though not necessarily in the correct geography. The remaining two statuses classified the enumeration as erroneous, either because of duplication or to reasons other than duplication.

Table 19: Unresolved Rates for Enumeration Statuses for Components of Census Coverage (Unweighted)

Enumeration Status for Components of Coverage	2020 Unresolved (Percent)	2010 Unresolved (Percent)
Correctly Enumerated		
(1) In the block search area	11.5	6.0
(2) In the correct county and place but outside of the block search area	12.1	6.7
(3) In the correct county and the wrong place	11.2	6.1
(4) In the wrong county and the correct place	3.9	2.1
(5) In the wrong county and place but the correct state	12.0	6.6
(6) In the wrong state	12.1	6.6
Erroneously Enumerated		
(7) Duplicates	0.4	0.3
(8) Erroneously enumerated for reasons other than duplication	11.5	6.0

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

In Table 19 we see that six of the eight enumeration statuses had unresolved rates of about 12 percent. Generally, these resulted from the same 12 percent of the enumerations. One enumeration status for which fewer enumerations were unresolved was Erroneously Enumerated because of Duplication, with about 0.4 percent unresolved. There was a much smaller amount of missing data here because only enumerations for which PES operations identified a possible duplicate with another census person were considered unresolved for the duplicate component. If we did not identify a possible duplicate link to another census record, then we assumed that no such duplicate record existed and treated the case as resolved to not be a duplicate. Another enumeration status for which there was less missing data was Enumerated in the Wrong County but the Correct Place, which had only a 3.9 percent unresolved rate. There were not as many blocks for which this status was possible.

Table 20 shows two estimated rates for each of the eight enumeration statuses: first without the imputation for unresolved enumeration statuses, that is, including only records with all enumeration statuses resolved, and then with the imputation for unresolved statuses. Unlike the results in Table 19, these rates are weighted. We see the largest effect of imputation for the status Correctly Enumerated in the Block Search Area, which decreased by about half a percentage point. The enumeration status In the Correct County and Place but Outside of the Block Search Area also saw an increase from 0.62 percent to 0.89 percent.

Table 20: Rates for Enumeration Statuses With and Without Imputation (Weighted)

Enumeration Status for Components of Coverage[1]	Resolved (Percent)	After Imputation (Percent)
Correctly Enumerated	97.74	97.71
(1) In the block search area[2]	96.48	95.86
(2) In the correct county and place but outside of the block search area	0.62	0.89
(3) In the correct county and the wrong place	0.21	0.31
(4) In the wrong county and the correct place	0.02	0.03
(5) In the wrong county and place but the correct state	0.20	0.30
(6) In the wrong state	0.22	0.31
Erroneously Enumerated	2.27	2.29
(7) Duplicates	1.65	1.66
(8) Erroneously enumerated for reasons other than duplication[2]	0.63	0.64

[1]This table does not include census whole-person imputations.
[2]The estimated rate of this status reflects the adjustment for duplication described in Section 4.2.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release).

6.4 The Cell Means Method to Impute for Enumerations Statuses for Components of Coverage

If we did not know which of the eight enumeration statuses should have been a "yes" for an enumeration, we assigned a probability for each of the eight enumeration statuses using a cell means imputation procedure. We created eight sets of cells, one for each enumeration status, and implemented the cell means method separately for each of the eight statuses. The number of cells varied by enumeration status because certain types of geography or because certain patterns of unresolved enumerations were impossible for a given status.

In many situations one or more of the statuses were not possible because of known geographic restrictions or because of other information collected about the enumeration. When one of the eight enumeration statuses was not possible for an enumeration with an unresolved status, we assigned a value of 0 to the impossible status.

In the cell means method, for each enumeration status, we put each resolved enumeration into one cell. For example, for the cell means imputation for the status Correctly Enumerated in the Block Search Area there were ten cells. Thus we put resolved enumerations into one of these ten cells. For the imputation for a given status, the predicted probability of that status for a given cell was the weighted proportion of that status among all of the cases with resolved statuses in the cell.

The basic formula for the calculation of the cell means probabilities was the same for all eight enumeration statuses. For a given enumeration status *i* and imputation cell *C*, each enumeration in cell *C* that was unresolved for status *i* was assigned the probability $P^*_{i,C}$. This probability was calculated with the following formula using only the enumerations in cell *C* that were resolved for enumeration status *i*:

$$P^*_{i,C} = \frac{\sum_{j \in C} W_j \times P_{i,j}}{\sum_{j \in C} W_j} \tag{7}$$

where,

- W_j is the original E-sample weight for the j th resolved enumeration in cell C.
- $P_{i,j}$ is the probability of the j th resolved enumeration in cell C being resolved as outcome i.

Note that the probabilities $P_{i,j}$ were 1 or 0. Then, for each enumeration status outcome *i*, every enumeration j was assigned a probability of being the *i*th outcome:

$$P_{i,j} = \begin{cases} 1, \text{ if enumeration } j \text{ is resolved as outcome } i \\ 0, \text{ if enumeration } j \text{ is resolved not as outcome } i \\ P^*_{i,C}, \text{ if enumeration } j \text{ is unresolved for outcome } i \end{cases} \tag{8}$$

where *C* represents the cell into which enumeration *j* was placed.

6.5 Adjustments to Cell Means Probabilities

We made two adjustments to the probabilities obtained from cell means model. The first adjustment accounted for duplication to census people who were subsampled out of the E sample and was similar to the adjustment for the correct enumeration status for net coverage error estimation in Section 4.2. The probability of the status Correctly Enumerated in the Block Search Area was decreased for the enumeration that had been coded a correct enumeration by PES. The same amount was added to the probability of status Erroneously Enumerated Because of Duplication.

The second adjustment controlled the sum of the eight enumeration status probabilities to equal 1 for each E-sample person enumeration. The imputation of each of the eight statuses was done independently, so this final adjustment was done to make sure the sum of the imputed probabilities equaled 1. In certain situations, the imputation method did not result in the enumeration status probabilities summing to 1. For each person enumeration *j*, for a particular enumeration status *i* , the adjusted probability $Pr_{adjusted,i,j}$ was the imputed probability divided by the sum of the eight enumeration status probabilities, as shown in the following equation.

$$Pr_{adjusted,i,j} = \frac{Pr_{i,j}}{\sum_{i=1}^{8} Pr_{i,j}} \tag{9}$$

6.6 Constructing the Imputation Cells

We formed eight sets of cells, one for each of the eight enumeration statuses. The resolved enumerations (those with no statuses missing) typically contributed to imputation cells for multiple statuses. The enumerations with unresolved statuses received imputed probabilities for each missing status from the set of cells corresponding to the missing status.

We constructed separate sets of cells for each enumeration status to assure that only permissible probabilities would be assigned to enumerations with unresolved statuses. The sample geography and the response data led to two key constraints that circumscribed the formation of the cells. We define these two constraints in later sections as the Type of Geography and the Pattern of Unresolved Statuses. We had separate cells for each enumeration status to accommodate the restrictions dictated by the Type of Geography and the Pattern of Unresolved Statuses, because these varied by the enumeration status.

In contrast to the 2020 PES methodology, the 2010 CCM used two sets of cells to impute the eight statuses. It then set equal to zero the probabilities for the impossible statuses for a given geography and rescaled the imputed probabilities so that their total equaled 1. However, the 2010 CCM did not take into account the responses (the pattern of unresolved statuses) and thus assigned positive probabilities to statuses that were impossible given the information collected in the interview.

6.6.1 Type of Geography

When constructing imputation cells there were two restrictions dictated to us by the data we collected. The first was the Type of Geography. When imputing each enumeration status, we considered whether the status was possible for the enumeration's sample geography, as some enumeration statuses were impossible for certain sample geographies. Consider several examples: in a place that was completely within a county, a person could not be counted in the correct place but a wrong county, enumeration status (4); in a place that was equivalent to a county, a person could not be counted in the correct county but a wrong place, enumeration status (3); an enumeration in Washington, DC, was not eligible for statuses (3), (4), or (5) because the place, county, and state all have the exact same boundary.

For situations where the geography of the sample enumeration made impossible one or more of the eight enumeration status outcomes, the probabilities of the non-applicable statuses were set to 0. For the definitions of the five types of geographies refer to Table 21. Most blocks were in Type 3 geographies; that is, they were located in a place fully within a county. In the Washington, DC, example, an enumeration with a missing enumeration status, would be assigned a 0 for statuses (3), (4), and (5), rather than imputing a positive probability for those statuses.

Table 21: Type of Geography and Restrictions

Definition	Example	Restricted or Impermissible Enumeration Statuses
All enumeration statuses possible	Chicago, IL; Atlanta, GA	None
State equivalent, county, and place are the same	District of Columbia (the only instance)	• Enumerated in the Correct County and a Wrong Place • Enumerated in a Wrong County and Place, but the Correct State • Enumerated in a Wrong County and the Correct Place
County inside Place	Queens County, in New York City	• Enumerated in the Correct County and a Wrong Place
Place inside County	City of Bowie, in Prince George's County, Maryland	• Enumerated in a Wrong County and the Correct Place
County and place are the same	City of Alexandria, Virginia	• Enumerated in the Correct county and a Wrong Place • Enumerated in a Wrong County and the Correct Place

Note: For tabulation purposes the District of Columbia was considered to be equivalent to a state.

6.6.2 Patterns of Unresolved Enumeration Statuses

The information that the PES follow-up interview collected led to the second restriction. The PES data collection yielded a limited number of possible combinations of unresolved enumeration statuses for components of coverage. The pattern of unresolved enumeration statuses refers to this combination of the eight statuses that were missing. In subsequent tables we name the variable that summarized the pattern of unresolved statuses "Unresolved Pattern."

The 2020 PES used person follow-up information to restrict the imputation of probabilities only to enumeration statuses that were compatible with the follow-up information. Some enumerations might have a missing value for all eight statuses, while other enumerations might have a missing value for two statuses (the other six statuses were resolved as "no"). Verbal descriptions of the eight possible combinations of unresolved statuses, labeled A through H, are shown in Table 22.

Table 22: Descriptions of Patterns of Unresolved Statuses

Unresolved Pattern	Description
A	Not in the sample block: In same county but unresolved if in same place as sample block
B	Not in the sample block: In the same state and place, but unresolved if in the same county as sample block
C	Not in the sample block: In the same state and a different place, but unresolved if in the same county as sample block
D	Not in the sample block: In different county but unresolved if in same place as sample block
E	Not in the sample block: Unresolved if in same county and if in same place as sample block
F	Not in the sample block: Unresolved if in same state as sample block
G	All enumeration statuses unresolved except for duplication
H	All enumeration statuses unresolved including duplication

Note that for Unresolved Patterns A – F, the interview data showed that the person was enumerated in the wrong block but had incomplete information about where they should have been counted. Such enumerations were considered resolved as a “no” for enumeration statuses 1, 7, and 8, and those probabilities set to 0. However, they could be unresolved for enumeration statuses 2, 3, 4, 5, and 6, depending on how much information we had about the address where they should have been counted.

Table 23 shows the enumeration statuses that were possible for each of the patterns of unresolved enumerations. Consider Unresolved Pattern A (described in Table 22): “Not in the sample block: In same county but unresolved if in same place as sample block.” Only two enumeration statuses were possible: (2) In the Correct County and Correct Place, but Outside the Block Search Area; and (3) In the Correct County but a Wrong Place. Note that the enumeration status Erroneously Enumerated because of Duplication (7) was unresolved only for the unresolved pattern H; for the others it was resolved as “no.”

Table 23: Possible Enumeration Status for a Given Pattern of Unresolved Statuses

Unresolved Pattern	Enumeration Status Possible?							
	1	2	3	4	5	6	7	8
A	No	Yes	Yes	No	No	No	No	No
B	No	Yes	No	Yes	No	No	No	No
C	No	No	Yes	No	Yes	No	No	No
D	No	No	No	Yes	Yes	No	No	No
E	No	Yes	Yes	Yes	Yes	No	No	No
F	No	Yes	Yes	Yes	Yes	Yes	No	No
G	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
H	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

6.6.3 Defining Imputation Cells

Each of the eight enumeration statuses had a separate set of imputation cells. Two of these are shown as examples in this section, and the other six in Attachment E.

The Unresolved Pattern specified which enumeration statuses were possible for a given unresolved case, and thus which resolved enumerations were eligible to be donors for that case. This was the key relationship that tied the resolved and unresolved enumerations together. For example, consider an enumeration in a county/place equivalent geography (Type of Geography = 4) without a Duplicate Link. There was only one possible Unresolved Pattern for such unresolved enumerations: ”Not a duplicate, but other enumeration statuses unresolved,” unresolved pattern ”G.” For these unresolved enumerations, five enumeration statuses were possible: (1) Correctly enumerated in the block search area; (2) Correctly enumerated in the right county and place but outside the block search area; (5) Correctly enumerated in a wrong county and wrong place, but in the correct state; (6) Correctly enumerated in a wrong state; and (8) Erroneously enumerated because of reasons other than duplication.

The cells for a given enumeration status accounted for all possible unresolved outcomes that could arise from combinations of Unresolved Pattern, Type of Geography, and Duplicate Link. For example, enumeration status (1), Correctly Enumerated in the Block Search area, had ten cells because there were ten combinations of Unresolved Pattern, Type of Geography, and Duplicate Link possible for this enumeration status. Refer to Table 24 in the following example for the definition of these cells. In contrast, enumeration status (2), Correctly Enumerated in the Correct County and Place but Outside of the Block Search Area, had 23 cells because there were 23 permissible combinations of these variables. Refer to Table E1 in Attachment E for the definitions of these 23 cells.

We present two examples to illustrate the construction of imputation cells. Table 24 shows the imputation cells for the enumeration status Correctly Enumerated in the Correct Block Search Area (1). Several points are worth noting.

- First G and H were the only patterns of unresolved enumeration statuses possible for the status Correctly Enumerated in the Block Search Area.
- Second, for each cell the resolved donors include enumerations with status (1), which contributed a positive probability; the rest of the donor enumerations in a given cell contributed zero to the cell probability.
- Third, the resolved donors that contributed to a cell depended on the cell's Type of Geography, the Duplicate Link Indicator, and the Unresolved Pattern.

Table 24: Cell Assignments for Enumeration Status (1): Correctly Enumerated in the Block Search Area

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	G	1, 2, 3, 4, 5, 6, 8
2		Yes	H	1, 2, 3, 4, 5, 6, 7, 8
3	1	No	G	1, 2, 6, 8
4		Yes	H	1, 2, 6, 7, 8
5	2	No	G	1, 2, 4, 5, 6, 8
6		Yes	H	1, 2, 4, 5, 6, 7, 8
7	3	No	G	1, 2, 3, 5, 6, 8
8		Yes	H	1, 2, 3, 5, 6, 7, 8
9	4	No	G	1, 2, 5, 6, 8
10		Yes	H	1, 2, 5, 6, 7, 8

A second example of a set of cells is seen in Table 25: Cell Assignments for Enumeration Status (6): Correctly Enumerated in a Wrong State.

Table 25: Cell Assignments for Enumeration Status (6): Correctly Enumerated in a Wrong State

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	F	2, 3, 4, 5, 6
2		No	G	1, 2, 3, 4, 5, 6, 8
3		Yes	H	1, 2, 3, 4, 5, 6, 7, 8
4	1	No	F	2, 6
5		No	G	1, 2, 6, 8
6		Yes	H	1, 2, 6, 7, 8
7	2	No	F	2, 4, 5, 6
8		No	G	1, 2, 4, 5, 6, 8
9		Yes	H	1, 2, 4, 5, 6, 7, 8
10	3	No	F	2, 3, 5, 6
11		No	G	1, 2, 3, 5, 6, 8
12		Yes	H	1, 2, 3, 5, 6, 7, 8
13	4	No	F	2, 5, 6
14		No	G	1, 2, 5, 6, 8
15		Yes	H	1, 2, 5, 6, 7, 8

7. References

Heim, K., “2020 Post-Enumeration Survey: Net Coverage Estimation,” DSSD 2020 Census Post-Enumeration Survey Memorandum Series #2020-J-07, U.S. Census Bureau, 2022.

Hogan, H., “The Accuracy and Coverage Evaluation: Theory and Design,” Survey Methodology, 29, U.S. Census Bureau, 2003, pp. 129–138.

Jost, M. and S. Konicki, “Post-Enumeration Survey Noninterview Adjustment Research Results,” DSSD 2020 Post-Enumeration Survey Memorandum Series #2020-E-04, U.S. Census Bureau, 2018.

Kennel, T., “The Design of the Post-Enumeration Survey for the 2020 Census,”
DSSD 2020 Post-Enumeration Survey Memorandum Series #2020-B-01, U.S. Census Bureau, 2019.

Khubba S., J. Hong, and K. Heim, “2020 Post-Enumeration Survey Estimation Report: Summary of Estimates of Coverage for People in the United States,” DSSD 2020 Census Coverage Measurement Memorandum Series #2020 G-01, U.S. Census Bureau, 2022.

Konicki, S., R. Bray, A. Keller, M. Seiss, and C. Viehdorfer, "2010 Census Coverage Measurement Estimation Methods: Missing Data," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-J-03, U.S. Census Bureau, 2013.

Marra, E., and T. Kennel, "Source and Accuracy of the 2020 Post-Enumeration Survey Person Estimates," DSSD 2020 Census Coverage Measurement Memorandum Series #2020 J-01, U.S. Census Bureau, 2022.

Phan, N., and J. Lawrence, "2020 Census Coverage Measurement Estimation Report: Characteristic Imputation Methods and Results," DSSD 2020 Post-Enumeration Survey Memorandum Series #2020-J-04, U.S. Census Bureau, 2022.

U.S. Census Bureau, Geographic Areas Reference Manual, 1994. https://www2.census.gov/geo/pdfs/reference/GARM/Ch9GARM.pdf

Valliant, R., J.A. Dever, and F. Kreuter, Practical Tools for Designing and Weighting Survey Samples, Springer Science+Business Media, New York, 2013.

Zamora, J., "2020 Post-Enumeration Survey Estimation Design," 2020 Post-Enumeration Survey Memorandum series #2020-J-03, U.S. Census Bureau, 2022.

Attachment A: Noninterview Adjustment Models

Table A1: Noninterview Adjustment Model

Model	Modeling Parameters		Stratification Parameters	
	Fixed Effects	Random Effect	Geography Strata	Number of Propensity Classes
Non-AIC	• State • Recoded type of address • Block sampling stratum code • Recode of Person Interview Reopen indicator • Recode of Initial Housing Unit Before Followup Match Code • Interaction between state and recode of Person Interview Reopen indicator	Block	State	5
AIC	• Recoded type of address • Recode of Initial Housing Unit Before Followup Match Code	Block	Subregion	5

Note: The subregions were the Northeast, South, and Midwest regions, and the West region split into two parts.

Attachment B: E-Sample Imputation Models

Table B1: Model Variables Used in Status Imputation for Correct Enumeration Status

Variable	Correct Enumeration Model 1	Correct Enumeration Model 2
Race/Origin Domain	X	X
Tenure	X	X
Sex	X	X
Age and Sex Group	X	X
Census Proxy Flag	X	X
Type of Census Response	X	X
Characteristic Imputation Flag	X	X
Relationship Type	X	X
BFU Match Code Group	X	X
Duplicate Link Flag	X	
Seasonal Address Flag	X	
Outmover Address Flag	X	
Inmover Address Flag	X	
Job Address Flag	X	
Military Address Flag	X	
Group Quarters Address Flag	X	
Relatives Address Flag	X	
College Address Flag	X	
Household with a Spousal Relationship	X	X
2010 CCM Correct Enumeration Rate by Tract	X	X
Relationship by Census Response Type Interaction	X	X
Relationship by Duplicate Link Flag Interaction	X	
Seasonal Address Flag by Duplicate Link Flag Interaction	X	

Attachment C: P-Sample Imputation Models

Table C4: Unresolved Rates for Inclusion, Mover, and Match Statuses: the 2020 PES and 2010 CCM Compared (Unweighted)

	2020				2010			
	Total	Resolved	Unresolved	Unresolved Rate (Percent)	Total	Resolved	Unresolved	Unresolved Rate (Percent)
All Cases								
P-Sample Inclusion Status	301,000	279,000	21,000	6.98	383,000	372,000	11,000	2.87
P-Sample Mover Status	284,000	264,000	19,500	6.87	356,000	345,000	11,000	3.09
P-Sample Match Status	284,000	261,000	22,000	7.75	356,000	343,000	13,000	3.65
P-Sample Inmover Match Status	39,000	16,500	22,000	56.41	37,500	24,500	13,000	34.67
P-Sample Nonmover Match Status	264,000	252,000	12,500	4.73	329,000	323,000	6,500	1.98
Sufficient Information Cases								
P-Sample Inclusion Status	288,000	279,000	8,700	3.02	377,000	372,000	4,800	1.27
P-Sample Mover Status	271,000	264,000	7,100	2.62	349,000	345,000	4,800	1.38
P-Sample Match Status	271,000	261,000	9,700	3.58	349,000	343,000	6,600	1.89
P-Sample Inmover Match Status	26,500	16,500	9,700	36.60	31,000	24,500	6,600	21.29
P-Sample Nonmover Match Status	252,000	252,000	80	0.03	323,000	323,000	50	0.02
Insufficient Information Cases								
Total	12,500	0	12,500	100	6,400	0	6,400	100

Note: Counts may not sum to totals shown because of rounding.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

Table C5: Model Variables Used in Status Imputation for P-Sample Inclusion and Match Status

Variable	Inclusion Status			Inmover Status		Nonmover Match Status		Inmover Match Status		
	Suff. 1	Suff. 2	Insuff.	Suff.	Insuff.	Suff.	Insuff.	Suff. 1	Suff. 2	Insuff.
Race/Origin Domain	X	X	X	X	X	X	X	X	X	X
Tenure	X	X	X	X	X	X	X	X	X	X
Correlation Bias Age-Sex Group	X	X	X	X	X	X	X	X	X	X
Proxy Flag	X	X	X			X	X	X	X	X
% of Pop Moving Past Year				X	X			X	X	X
Recoded Type of Address	X	X	X			X	X	X	X	X
Char. Imp. Flag	X	X	X			X	X	X	X	
Relationship Type	X	X	X					X	X	X
Edited Relationship Code				X	X	X	X			
Recoded Roster Flag	X	X	X			X	X			
BFU Match Code Group	X	X								
Collapsed BFU Match Code Group						X				
Alternative BFU Match Code Group								X		
BFU Insuff. Info Group			X		X					
BFU Insuff. Info Group 2							X			
Second Alternative BFU Insuff. Info Group										X
2010 CCM Tract-Level Person Match Rate						X	X	X	X	X
State						X	X			
American Indian Country Indicator								X	X	X
Spousal Household						X	X	X	X	X
Wave						X	X	X	X	X
Attempt Type Recode								X	X	X
MSATEA Group								X	X	X
Inmover Address Flag	X									
Seasonal Address Flag	X									
Other Address Flag	X					X				
Followup Flag				X					X	
Ages 0-24 College Flag	X									
Followup Flag*% of Pop Moved in Past Year				X						
BFU Insuff. Info Group *% of Pop Moving Past Year					X					
Tenure*Correlation Bias Age-Sex Group				X	X					
Tenure*Domain								X	X	X

Attachment D: Variable Descriptions

Table D1: Model Variable Descriptions

Variable Description	Variable Name	Valid Values
State	STATE	01: Alabama 02: Alaska ⋮ 56: Wyoming
Recoded Type of Address	HUTOA2	1: Single family 2: Multi unit address 3: Other
Block Sampling Stratum Code	SS1	1: Small (0-2 HUs) 2: Medium renter (3-57 HUs, over specified percent rented) 3: Medium owner (3-57 HUs, over specified percent owned) 4: Large renter (58+ HUs, over specified percent rented) 5: Large owner (58+ HUs, over specified percent owned) 6: American Indian Reservation strata
Recode of Person Interview Reopen Indicator	WAVE_CLASS2	0: Interviews, vacant, and nonexistent in the original Person Interview and the reclassified noninterviews in the original Person Interview and Person Interview Reopen 1: Interviews, vacant, and nonexistent in the Person Interview Reopen. The no-contact noninterviews in the original Person Interview and Person Interview Reopen.
Recode of Initial Housing Unit Before Followup Match Code	HUBFUMC_CLASS	M: Match P: Possible match X: Other
Interaction between State and WAVE_CLASS2	STATE*WAVE_CLASS2	01 * 0: Alabama in WAVE_CLASS2 category 0
Race/Hispanic Origin Domains	DOMAIN	1: American Indian/Alaska Native On Reservations 2: American Indian/Alaska Native Off Reservations 3: Hispanic 4: Non-Hispanic Black 5: Native Hawaiian or Pacific Islander 6: Asian 7: White or Some Other Race
Occupancy/Tenure of the P-sample Person	OCCTEN	1: Owner 2: Renter

Variable Description	Variable Name	Valid Values
Age/Sex Group	CBAGESEX	01: 0-4 02: 5-9 03: 10-17 04: 18-24 Male 05: 18-24 Female 06: 25-29 Male 07: 25-29 Female 08: 30-49 Male 09: 30-49 Female 10: 50-64 Male 11: 50-64 Female 12: 65+ Male 13: 65+ Female
Person Interview Proxy Flag	PI_PROXY_FLAG	0: Person Interview Conducted with a Household Respondent 1: Person Interview Conducted with a Proxy Respondent 2: Person Record in a Housing Unit Assigned by the Noninterview Adjustment Procedure
P-Sample Characteristic Imputation Flag	CHAR_IMP_FLAG	0: All Characteristics Reported 1: At Least One Characteristic Imputed
Relationship Type	REL_TYPE	1: Nuclear Family Member 2: Adult Child of the Householder 3: Other Member of the Household
Edited Relationship Code	RELSHIP	20: Householder 21: Opposite-Sex Husband/Wife/Spouse 22: Opposite-Sex Unmarried Partner 23: Same-Sex Husband/Wife/Spouse 24: Same-Sex Unmarried Partner 25: Biological Son/Daughter 26: Adopted Son/Daughter 27: Stepson/Stepdaughter 28: Brother/Sister 29: Father/Mother 30: Grandchild 31: Parent-in-Law 32: Son-in-Law/Daughter-in-Law 33: Other Relative 34: Roommate or Housemate 35: Foster Child 36: Other Nonrelative

Variable Description	Variable Name	Valid Values
Recoded Roster Flag	CM_ROSFLG	2: Initial Roster 3: Person Stays Here Often Probe 4: Person Trying to Find a Home Probe 5: Baby Probe 6: Other Relative Probe 7: Added at Review list 8: Outmover and Outmover Review list 10: Whole Household Outmover
P-sample Before Followup Match Code Group	PBFUMCG	Nonfollowup Cases 01: Resolved Before Followup – Nonmover 02: Resolved Before Followup – Inmover 03: Unresolved BFU, Resolved AFU 04: Delete BFU Followup Cases 05: Possible Matches 06: Conflicting Household 07: Nonmover, Whole Household Nonmatch 08: Nonmover, Partial Household Hon match 09: Inmover with Ungeocoded Address or Nonmatch Person – Before Followup Inmover 10: Possible Duplicates or Unclassified Residence Status 11: Other PFU Types 12: Not a Housing Unit in Initial Housing Unit Followup 13: Insufficient Information for Matching
Collapsed P-sample Before Followup Match Code Group	PBFUMCG2	Nonfollowup Cases 01: Resolved Before Followup – Nonmover 05: Possible Matches 06: Conflicting Household 08: Nonmover – Partial Household Nonmatch 10: Possible Duplicates or Unclassified Residence Status 11: Other PFU types 13: Insufficient Information for Matching All others: 99: All cases with PBFUMCG in 2, 3, 4, 7, 9, or 12

Variable Description	Variable Name	Valid Values
Alternative BFU Match Code Group	PBFUMCG_ALT	Nonfollowup Cases 01: Resolved Before Followup – Nonmover 02: Resolved Before Followup – Inmover/Unresolved BFU, Resolved AFU 04: Delete BFU Followup Cases 05: Possible Matches 06: Conflicting Household 07: Nonmover, Whole Household Nonmatch 08: Nonmover, Partial Household Nonmatch 09: Inmover with Ungeocoded Address or Nonmatch Person – Before Followup Inmover 10: Possible Duplicates or Unclassified Residence Status/ Not a Housing Unit in Initial Housing Unit Followup 11: Other PFU Types 13: Insufficient Information for Matching
P-sample Before Followup Insufficient Information Group	PBFUKIG	1: Inmover 2: Unresolved Inclusion Status 3: HU Matched and Not a Conflicting Housing Unit 4: HU Not Matched and Not a Conflicting Housing Unit 5: Conflicting Household 6: Delete BFU
P-sample BFU Insufficient Information Group 2	PBFUKIG2	BFU Nonmovers or Insufficient Information cases 1: Nonmover or KI in a Matched and Not a Conflicting Housing Unit 2: Nonmover or KI in a Not Matched and Not a Conflicting Housing Unit 3: Nonmover or KI in a Conflicting Household All others 4: Sufficient information Case that is Not a Nonmover
Second Alternative BFU Insuff. Info Group	PBFUKIG_ALT2	1: Inmover 2: Unresolved Inclusion Status 3: HU Matched and Not a Conflicting Housing Unit 4: HU Not Matched and Not a Conflicting Housing Unit 5: Conflicting Household/Delete BFU
Final Inmover Address Flag	FINIMVFLG	0: No Inmover Address Currently Attached to the PI Person or Linked Census Person 1: Inmover Address Currently Attached to the PI person or Linked Census Person
Final Seasonal Address Flag	FINSEAFLG	0: No Seasonal Address Currently Attached to the PI Person or Linked Census Person 1: Seasonal Address Currently Attached to the PI Person or Linked Census Person

Variable Description	Variable Name	Valid Values
Has Outmover, College, Relative, Military, Job, or GQ Address	OTHADDFLG	0: No Other Address Flags 1: One or More of the Following Address Flags Attached to Person: Outmover, College, Relative, Military, Job, Group Quarters
Ages 0-24 College Flag	COL_FLG	0: Person is 25 or older, or a College Address is not Currently Attached to their PES record or Linked Census Enumeration 1: Person is Less than 25 and a College Address is Currently Attached to their PES record or Linked Census Enumeration
Followup Indicator Flag	FU_FLAG	Y: Followup is Needed N: Followup is Not Needed
% of Pop Moving Past Year	PCT_DIFF_HU_1YR_AGO_ ACS_15_19	[Continuous Variable]
2010 CCM Tract-Level Person Match Rate	CCM_PER_MAT_RATE_TR	[Continuous Variable]
State	BCUSTATEFP	[Refer to State FIPS Codes]
American Indian Country Indicator	AICIND	0: Not on Indian Country 1: American Indian Reservation 2: Indian Country Off American Indian Reservation
Spousal Household	SPOUSAL	0: No Spouse in Household 1: Spouse Present in Household
MSATEA Group	MSATEA	0: Large MSA Self-Response 1: Medium MSA Self-Response 2: Small MSA Self-Response 3: Non-MSA Self-Response 4: Large, Medium, or Small MSA Update/Leave 5: Non-MSA Update/Leave or Update/Enumerate
Wave	WAVE	1: Person Interview Conducted During First Wave 2: Person Interview Conducted During Second Wave 3: Person Interview Conducted During Third Wave 4: Person Interview Conducted During Fourth Wave (PI Reopen)
Attempt Type Recode	ATTEMPT_CLASS2	1: Person Visit to Sample Address 2: Person Visit to Proxy Respondent 3: All Other Attempts

Variable Description	Variable Name	Valid Values
Census Proxy Flag Collapsed	RCENPROX2	Blank H: Household member on April 1 or Household member moved in after April 1 O: Other (multiple respondent types) P: Neighbor or other proxy respondent
CCM Correct Enumeration Rate by Tract	CCM_PER_CE_RATE_TR	Continuous
Session Context Code	SCCG2	1: Internet Self-Response 2: Paper Questionnaire Self-Response 3: Electronic Enumeration or Paper Enumeration 5: Administrative Records 8: Coverage Followup
E-sample BFU Match Code Group	EBFUMCG3	1: Resolved Before Followup 2: Possible Matches 3: Conflicting Household 4: Partial Household Nonmatch 5: Whole Household Nonmatch 6: Unresolved Inclusion Status 7: Duplicate 9: Insufficient Information for Dual System Estimation
Duplicate Link Flag	DUPLINK_IND	0: No duplicate link attached to person 1: Duplicate link attached to person
E-sample Seasonal Address Flag	PAM_FI_SEAS	0: No seasonal address attached to person 1: Seasonal address attached to person
E-sample Outmover Address Flag	PAM_FI_OUTMVR	0: No outmover address attached to person 1: Outmover address attached to person
E-sample Inmover Address Flag	PAM_FI_INMVR	0: No inmover address attached to person 1: Inmover address attached to person
E-sample Job Address Flag	PAM_FI_JOB	0: No Job address attached to person 1: Job address attached to person
E-sample Military Address Flag	PAM_FI_MIL	0: No Military address attached to person 1: Military address attached to person
E-sample Group Quarters Address Flag	PAM_FI_GQ	0: No Group Quarters address attached to person 1: Group Quarters address attached to person

Variable Description	Variable Name	Valid Values
E-sample Relative Address Flag	PAM_FI_REL	0: No Relative address attached to person 1: Relative address attached to person
E-sample College Address Flag	PAM_FI_COLL	0: No College address attached to person 1: College address attached to person

Note: The initial housing unit match codes before follow-up is the match codes assigned during the Initial Housing Unit (IHU) operation. For more information on IHU operations refer to Kennel 2019.

Attachment E: Cells Used in Imputation of Person Components of Census Coverage for the U.S.

Table E1: Cell Assignments for Enumeration Status (2): Correctly Enumerated in the Correct County and Place, but Outside of the Block Search Area

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	A	2, 3
2		No	B	2, 4
3		No	E	2, 3, 4, 5
4		No	F	2, 3, 4, 5, 6
5		No	G	1, 2, 3, 4, 5, 6, 8
6		Yes	H	1, 2, 3, 4, 5, 6, 7, 8
7	1	No	F	2, 6
8		No	G	1, 2, 6, 8
9		Yes	H	1, 2, 6, 7, 8
10	2	No	B	2, 4
11		No	E	2, 4, 5
12		No	F	2, 4, 5, 6
13		No	G	1, 2, 4, 5, 6, 8
14		Yes	H	1, 2, 4, 5, 6, 7, 8
15	3	No	A	2, 3
16		No	E	2, 3, 5
17		No	F	2, 3, 5, 6
18		No	G	1, 2, 3, 5, 6, 8
19		Yes	H	1, 2, 3, 5, 6, 7, 8
20	4	No	E	2, 5
21		No	F	2, 5, 6
22		No	G	1, 2, 5, 6, 8
23		Yes	H	1, 2, 5, 6, 7, 8

Table E2: Cell Assignments for Enumeration Status (3): Correctly Enumerated in the Correct County and a Wrong Place

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	A	2, 3
2		No	C	3, 5
3		No	E	2, 3, 4, 5
4		No	F	2, 3, 4, 5, 6
5		No	G	1, 2, 3, 4, 5, 6, 8
6		Yes	H	1, 2, 3, 4, 5, 6, 7, 8
7	3	No	A	2, 3
8		No	C	3, 5
9		No	E	2, 3, 5
10		No	F	2, 3, 5, 6
11		No	G	1, 2, 3, 5, 6, 8
12		Yes	H	1, 2, 3, 5, 6, 7, 8

Table E3: Cell Assignments for Enumeration Status (4): for Correctly Enumerated in a Wrong County and the Correct Place (4)

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	B	2, 4
2		No	D	4, 5
3		No	E	2, 3, 4, 5
4		No	F	2, 3, 4, 5, 6
5		No	G	1, 2, 3, 4, 5, 6, 8
6		Yes	H	1, 2, 3, 4, 5, 6, 7, 8
7	2	No	B	2, 4
8		No	D	4, 5
9		No	E	2, 4, 5
10		No	F	2, 4, 5, 6
11		No	G	1, 2, 4, 5, 6, 8
12		Yes	H	1, 2, 4, 5, 6, 7, 8

Table E4: Cell Assignments for Enumeration Status (5): Correctly Enumerated in a Wrong County and Wrong Place, but the Correct State

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	C	3, 5
2		No	D	4, 5
3		No	E	2, 3, 4, 5
4		No	F	2, 3, 4, 5, 6
5		No	G	1, 2, 3, 4, 5, 6, 8
6		Yes	H	1, 2, 3, 4, 5, 6, 7, 8
7	2	No	D	4, 5
8		No	E	2, 4, 5
9		No	F	2, 4, 5, 6
10		No	G	1, 2, 4, 5, 6, 8
11		Yes	H	1, 2, 4, 5, 6, 7, 8
12	3	No	C	3, 5
13		No	E	2, 3, 5
14		No	F	2, 3, 5, 6
15		No	G	1, 2, 3, 5, 6, 8
16		Yes	H	1, 2, 3, 5, 6, 7, 8
17	4	No	E	2, 5
18		No	F	2, 5, 6
19		No	G	1, 2, 5, 6, 8
20		Yes	H	1, 2, 5, 6, 7, 8

Table E5: Cell Assignments for Enumeration Status (7): Erroneously Enumerated Because of Duplication

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	Yes	H	1, 2, 3, 4, 5, 6, 7, 8
2	1	Yes	H	1, 2, 6, 7, 8
3	2	Yes	H	1, 2, 4, 5, 6, 7, 8
4	3	Yes	H	1, 2, 3, 5, 6, 7, 8
5	4	Yes	H	1, 2, 5, 6, 7, 8

Table E6: Cell Assignments for Enumeration Status (8): Erroneously Enumerated Because of Reasons Other than Duplication

Cell Number	Type of Geography	Duplicate Link	Unresolved Pattern	Donors - Resolved Enumeration Statuses
1	0	No	G	1, 2, 3, 4, 5, 6, 8
2	0	Yes	H	1, 2, 3, 4, 5, 6, 7, 8
3	1	No	G	1, 2, 6, 8
4	1	Yes	H	1, 2, 6, 7, 8
5	2	No	G	1, 2, 4, 5, 6, 8
6	2	Yes	H	1, 2, 4, 5, 6, 7, 8
7	3	No	G	1, 2, 3, 5, 6, 8
8	3	Yes	H	1, 2, 3, 5, 6, 7, 8
9	4	No	G	1, 2, 5, 6, 8
10	4	Yes	H	1, 2, 5, 6, 7, 8